CAPÍTULO 9

Significado e Implicações dos Resultados

Um pouco à maneira dos primeiros automóveis movidos a motor, que inutilmente conservavam o formato e as características das carruagens puxadas por cavalos, algumas das primeiras propostas de avaliação da inteligência no quadro de um novo paradigma de investigação, o sistémico, conservam em larga medida as características das técnicas diferenciais tradicionais. Seja pela inércia inerente a qualquer mudança súbita de movimento ou direcção, seja por motivos de ordem comercial[1], a verdade é que a estrutura, apresentação, forma de aplicação, tipos de conteúdos e até mesmo tipos de tarefas que constituem o *STAT-R (H*), a única técnica de medida sistémica da inteligência até hoje construída, são muito mais semelhantes aos tradicionais do que o autor leva a supor, a partir da sua persistente, veemente e até contundente crítica à avaliação clássica da inteligência humana. Uma técnica tradicional de avaliação da inteligência, como a *WAIS-III*, pelo contrário, embora concebida e construída no quadro do paradigma diferencial, apresenta algumas potencialidades para a avaliação sistémica da inteligência, quer pela riqueza da leitura complexa e contextual do desempenho de cada indivíduo, viabilizada pela aplicação individual, quer pela perspectiva estrutural em que radica, a qual reconhece e respeita níveis distintos de observação e explicação que não se substituem, antes se

[1] As negociações entre R. Sternberg e a *Psychological Corporation*, hoje parte da *Harcourt Assessment* (a mesma empresa que edita as escalas de Wechsler) tendo em vista a publicação do *STAT* tiveram início pouco tempo após o início dos trabalhos experimentais com a primeira versão, embora não tenham chegado a ser concluídas (comunicação pessoal).

articulam, se interligam de forma única em cada indivíduo, em função do desenvolvimento, da experiência e de factores conativos e afectivos tidos por inseparáveis do funcionamento cognitivo.

O presente capítulo pretende discutir os resultados obtidos com as duas técnicas de avaliação da inteligência, que se supõe representarem os dois paradigmas de investigação e avaliação da inteligência sob estudo, e retirar algumas implicações, do ponto de vista do desenvolvimento de metodologias de avaliação e investigação da inteligência, bem como do ponto de vista do seu significado nos planos teórico e metateórico. As questões de significado das medidas prefiguram um entendimento amplo e dialéctico da validação, cuja inspiração na concepção de validade de Messick (1975, 1980) é evidente, entendimento de que decorre a interpretação dos resultados desta investigação na dupla acepção da validação das medidas e da validação dos modelos teóricos subjacentes às medidas (Afonso, 1987a,b).

9.1. IMPLICAÇÕES METODOLÓGICAS DOS RESULTADOS

O primeiro nível em que se procede a uma reflexão acerca do significado e das implicações dos resultados é de natureza metodológica e assenta na discussão dos estudos metrológicos das duas técnicas diferenciais, *WAIS-III* e *STAT-R (H)*. Este estudo metrológico, ainda que acessório aos objectivos centrais da presente investigação, acabou por ocupar parte substancial do tempo e do esforço que lhe foi destinado, tendo mesmo contribuído para demorar a conclusão do projecto[2], e reflectiu-se também na extensão que ocupa no texto desta dissertação. O estudo metrológico era, contudo, incontornável, posto que ambas as técnicas foram traduzidas e adaptadas no âmbito do projecto, e apenas poderiam representar os dois paradigmas em análise se demonstrassem correspondência aos modelos teóricos que procuram operacionalizar.

Importa, pois, começar por traçar uma apreciação global dos resultados metrológicos para dela retirar algumas implicações quanto às possibilidades de aperfeiçoamento futuro das técnicas de medida ou, num nível mais fundamental, quanto à própria legitimidade da sua utilização no quadro dos respectivos paradigmas de investigação e avaliação. Parte-se, assim, de uma síntese dos resultados obtidos com cada instrumento, para proceder de seguida à identificação dos principais aperfeiçoamentos que sugerem e à discussão das suas implicações do ponto de vista da validade de construto das medidas.

[2] Ainda assim, o prazo estabelecido pelo *Centro de Psicometria e Psicologia da Educação* (*CPPE*) para a conclusão do projecto, coincidente, aliás, com o prazo contratualmente concedido pela *Harcourt Assessment* para direito de utilização da *WAIS-III* nesta investigação – 31 de Dezembro de 2006 – foi cumprido, uma vez que nessa data se encontrava concluído todo o estudo experimental, e já em fase avançada a redacção da presente dissertação.

9.1.1. *WAIS-III* (2002): Síntese e Significado Metrológico dos Resultados

Recorde-se que o estudo da versão experimental portuguesa da *WAIS-III* se baseou num único Ensaio Experimental, embora tenham sido conduzidos outros estudos, incidindo no aperfeiçoamento dos critérios de classificação e na estimação da precisão – estabilidade temporal dos resultados e acordo inter-avaliadores. A implementação de um único Ensaio Experimental justificou-se, por um lado, pelo bom nível metrológico global dos resultados, obtidos nesse Ensaio e nos outros estudos, e, por outro lado, pelo maior investimento que desde cedo se reconheceu necessário dedicar ao aperfeiçoamento do *STAT-R (H)*.

Resumem-se e discutem-se de seguida as principais conclusões que puderam ser extraídas dos estudos metrológicos da *WAIS-III*:

- A análise da dificuldade dos itens pôs em evidência um padrão de resultados coerente com as características das amostras (baixo nível etário e elevado nível de escolaridade em relação à população geral): percentagem de acertos relativamente elevada em todos os subtestes; médias dos sujeitos sistematicamente superiores às médias dos itens, quando medidos numa escala única (*logit*), no quadro do Modelo de Crédito Parcial (TRI) (resultados bem explícitos nos mapas dos sujeitos e dos itens apresentados em anexo); taxas de sucesso (que tomam em conta a gradação de pontuação nos subtestes que admitem crédito parcial) elevadas. A comparação entre níveis de escolaridade confirmou, aliás, existirem diferenças significativas favoráveis aos grupos de mais elevada escolaridade e a comparação entre grupos etários evidenciou tendência para resultados mais elevados no grupo mais jovem, à excepção do resultado compósito de Compreensão Verbal.
- Embora este conjunto de resultados seja, assim, coerente com as características das amostras, convém assinalar que em parte poderá ter decorrido de alguma contaminação introduzida pelos critérios de classificação, aperfeiçoados e testados com base na variável escolaridade; ainda assim, este estudo dos critérios incidiu apenas nos subtestes de resposta aberta – Vocabulário, Semelhanças e Compreensão – que não são os únicos em que se observa tendência para a superioridade dos grupos de escolaridade mais elevada. Conclui-se, por consequência, favoravelmente em relação à adequação global do nível de dificuldade da prova para a população geral portuguesa, sem contudo deixar de

assinalar a necessidade de novos estudos com amostras representativas da população.

- A ordem de dificuldade dos itens, tomando em conta os coeficientes de correlação entre as ordenações original e observada na amostra do Ensaio Experimental, sugere a necessidade de reordenação, sobretudo nos subtestes de Informação, Completamento de Gravuras, Semelhanças e Vocabulário, ou ainda, incluindo também os dados do Estudo Principal, nos subtestes de Cubos e Disposição de Gravuras. Considerando que a empresa editora da *WAIS-III* por princípio não autoriza a introdução de modificações dos subtestes de Realização, a eventual revisão quanto à ordenação da dificuldade deverá concentrar-se nos subtestes de Informação, Semelhanças e Vocabulário.
- Uma tal revisão não pode deixar de dar atenção quer aos índices de discriminação dos itens que, embora se situem em nível adequado quando tomados em conjunto, permitem detectar alguns casos de menor adequação, quer aos índices de ajustamento ao Modelo de Crédito Parcial, que em toda a prova assinalam menos de duas dezenas de itens desajustados, poucos deles com nível de desajustamento expressivo, e alguns compreensivelmente desajustados, por serem itens iniciais de cada subteste[3].
- O estudo da sensibilidade diferencial dos itens à variável sexo revelou ainda um número bastante significativo de itens favoráveis ao sexo masculino, de onde decorreram diferenças estatisticamente significativas encontradas, quer no Ensaio Experimental quer no Estudo Principal, em favor do sexo masculino. Uma revisão dos itens deverá atender também a estes dados.
- A apreciação global dos índices de precisão legitima a utilização da prova em investigação, podendo antecipar-se que a correcção de algumas deficiências ao nível dos itens venha a contribuir para melhorar a consistência interna. Os índices de bipartição e alfa de Cronbach situaram-se em geral acima de .80 e nalguns subtestes atingiram valores na ordem dos .90, embora tenham sido ligeiramente inferiores no Estudo Principal, em comparação com o Ensaio Experimental. O facto de o número de itens ser mais elevado no Estudo Principal, em que foram aplicados todos os itens sem exclusão dos de inversão como no Ensaio Experimental, não contribuiu, como se poderia esperar, para aumentar a

[3] Por uma questão de síntese e porque, por razões contratuais anteriormente aludidas, não pode ser revelado o conteúdo dos itens da *WAIS-III*, não é apresentada a lista dos itens que se identificaram passíveis de beneficiar de revisão.

consistência interna, possivelmente porque, dado o nível relativamente elevado de competência das amostras (parcialmente explicado pelo nível de escolaridade), os itens acrescentados (de inversão) se mostraram pouco discriminativos e pouco correlacionados com os restantes itens de cada subteste (por outras palavras, diminuíram o ajustamento ou calibração da dificuldade dos itens ao nível de competência da amostra). Na análise da consistência interna na amostra do estudo de estabilidade temporal e inter-avaliadores, os coeficientes Alfa de Cronbach corroboraram os que foram obtidos nas amostras do Ensaio Experimental e do Estudo Principal situando-se todos acima de .70, o critério considerado mínimo para utilização em investigação. Ainda assim, os subtestes que sugerem alguma atenção, do ponto de vista da consistência interna, são os de Composição de Objectos, Completamento de Gravuras, Disposição de Gravuras e Compreensão. O primeiro é já tradicionalmente um subteste com deficiente consistência interna (possivelmente devido ao reduzido número de itens), razão por que é hoje um subteste opcional no seio da *WAIS-III*, e razão também por que foi abandonado no Estudo Principal; dos restantes, apenas o de Compreensão poderá ser melhorado, posto não ser autorizada a modificação dos subtestes de Realização.

- Ainda no que toca ao estudo da precisão, os resultados, quer do estudo de estabilidade temporal quer do estudo de acordo inter-avaliadores, foram francamente positivos e situaram-se ao nível, ou até ligeiramente acima, dos resultados apurados nos estudos com a forma original do teste. A este respeito, importa notar que os critérios de classificação, após aperfeiçoamento e teste empírico no âmbito do estudo específico conduzido com esse objectivo, permitiram a obtenção de elevado acordo entre avaliadores experientes, e até mesmo entre avaliadores com menor experiência de classificação de escalas de Wechsler. Acresce que o estudo das categorias de cotação dos subtestes de Vocabulário, Semelhanças e Compreensão, no âmbito do Modelo de Crédito Parcial (TRI), demonstraram que apenas um reduzido número de itens (seis, no conjunto dos três subtestes) apresentou algum grau de desajustamento ou de desordenação das categorias de classificação, a que, ainda assim, importa atender aquando da revisão dos itens dos subtestes.
- O estudo da estrutura interna das medidas forneceu resultados muito semelhantes nas amostras em que foi efectuado: identificação de um factor em que saturam de forma muito expressiva todos os subtestes, antes da rotação; identificação de dois

factores com lambda superior a 1.00, um definido a partir dos subtestes Verbais e outro a partir dos subtestes de Realização (embora a importância relativa dos dois factores tenha sido diversa em diferentes amostras); quando retido o terceiro factor, já com valor próprio inferior à unidade e explicando, por consequência, uma parcela menor da variância total dos resultados, são os subtestes de Memória de Trabalho – Aritmética, Memória de Dígitos (e Ordenação de Letras e Números) – que saturam nesse factor, esboçando-se, por consequência, a aproximação da estrutura interna das medidas à estrutura tradicionalmente identificada nas escalas de Wechsler – estrutura que foi ainda a melhor ajustada no estudo factorial confirmatório; e quando retido o quarto factor, embora reconhecendo a sua pouca relevância na explicação da variância total, a estrutura aproxima-se da que corresponde aos índices factoriais da versão original da prova, pois nele saturam os subtestes de Velocidade de Processamento (Código:Dígito-Símbolo e Pesquisa de Símbolos). De entre os índices factoriais, o de Velocidade de Processamento é, no entanto, o que de forma sistemática menos claramente emerge nas amostras desta investigação, pois explica uma parte diminuta da variância dos resultados, possivelmente fruto da idade predominantemente jovem e homogénea das amostras estudadas (uma vez que a investigação demonstra que a velocidade de processamento declina com a idade – Afonso, 2004f, 2005d, 2006b).

- O estudo da dimensionalidade confirma a elevada percentagem de variância que é explicada pelo modelo, em cada subteste e em cada resultado compósito, e a reduzida importância do primeiro contraste, quando comparado com a medida do modelo. Particularmente importante é a verificação de que o primeiro contraste, no estudo da unidimensionalidade da Escala Completa, não é significativo, pois explica uma fracção mínima da variância total, quando comparado com a medida do modelo, embora uma breve inspecção dos itens que definem esse contraste ponha em evidência a oposição entre itens Verbais (Vocabulário e Informação) e itens de Realização (principalmente, Matrizes e Cubos), itens ajustados ao modelo, isto é, que não definem outras dimensões paralelas à do modelo, antes apoiam a subdivisão da prova (em parte Verbal e parte de Realização) ao mesmo tempo que reportam os seus resultados aos factores de inteligência cristalizada e inteligência fluida, respectivamente, presentes nos mais consensuais modelos factoriais das condutas cognitivas.

- Por fim, na análise conjunta da *WAIS-III* e do *STAT-R (H),* quer nos estudos factoriais, exploratório e confirmatório, quer no estudo da dimensionalidade, emergiu uma configuração das medidas compatível com a concepção de inteligência global de Wechsler, mais do que com a Teoria Triárquica de Sternberg: a emergência de um factor comum a todas as variáveis, quer na análise dos resultados totais, quer na análise dos resultados compósitos, e o estabelecimento de uma estrutura que se organiza em função de factores definidos a partir da *WAIS-III* (Realização/Verbal), mais do que do *STAT-R (H)* (domínios de processamento, ou mesmo áreas de conteúdo).

O conjunto de resultados obtido nos estudos metrológicos da *WAIS-III,* conduzidos no âmbito da presente investigação, constituiu sólido fundamento para a adopção desta técnica como representante do paradigma diferencial de avaliação da inteligência, e em particular como operacionalização do conceito de inteligência global definido por Wechsler: os diversos subtestes (e os resultados compósitos parciais) podem legitimamente ser tratados como medidas diversas de uma única inteligência e o resultado total da bateria representa a sua melhor medida isolada[4]. Convém que se refira que na comparação com resultados obtidos internacionalmente, quer com a versão original da *WAIS-III* (Wechsler, 1997b; ver também Ward, Ryan & Axelrod, 2000; Taub, McGrew & Witta, 2004) quer com as adaptações espanhola (Wechsler, 1999b) e francesa (Wechsler, 2000), a versão experimental portuguesa apresenta características metrológicas de nível equivalente, apesar do carácter assumidamente provisório e do reconhecimento da necessidade de aperfeiçoamento do conteúdo de vários subtestes.

9.1.2. *STAT-R (H)* (2002-2004): Síntese e Significado Metrológico dos Resultados

O estudo metrológico do *STAT-R (H)* ocupou a parte mais volumosa da presente investigação, uma vez que tornou necessárias múltiplas revisões e tentativas de aperfeiçoamento. Por se tratar de uma nova técnica de medição da inteligência, enquadrada por uma teoria ela própria em elaboração e sob avaliação empírica, e baseada em novos racionais de construção, esperava-se à partida que esta prova viesse a necessitar de enorme cuidado de adaptação e estudo prévio. Mesmo assim, estas expectativas foram francamente excedidas, posto que houve que desenvolver mais estudos empíricos prévios do que os inicialmente previstos. Assim, para além do Estudo Piloto, do qual decorreu uma primeira revisão, e do Ensaio Experimental inicialmente planeado, tornou-se necessário, após revisão,

[4] Assinale-se, a propósito, que o subteste de Vocabulário deve merecer uma especial atenção de revisão, atendendo a que apresentou um padrão de resultados discrepante do que é habitual nas escalas de Wechsler, em geral, e na *WAIS* em particular: não apresentou, por exemplo, os índices de consistência interna mais elevados, nem as mais altas saturações em *g*, nos diversos estudos.

um 2º Ensaio Experimental, e ainda um Ensaio Sem Tempo Limite, tendo em vista a aplicação de metodologia baseada nos modelos de Traço Latente.

Resumem-se de seguida os resultados obtidos nestes estudos, bem como na análise metrológica efectuada no âmbito do Estudo Principal:

- Numa síntese dos múltiplos estudos do *STAT-R (H)* efectuados nesta investigação, não se pode deixar de começar por reconhecer a fragilidade geral dos resultados metrológicos, desde a análise de itens ao estudos da precisão, desde a análise de Rasch ao estudo da estrutura interna das medidas. Mesmo assim, há a assinalar alguma evolução positiva, embora modesta, ao longo dos vários estudos, presumivelmente decorrente do sucessivo esforço de revisão da prova, embora parte da melhoria possa ser eventualmente atribuída a flutuações amostrais.
- Na análise de itens, a tendência foi no sentido de se revelarem grandes discrepâncias no nível de dificuldade, entre as várias partes, a par de baixos índices de discriminação, com frequência em partes da prova com elevadas percentagens de sucesso – Parte 1 (Analítica/Verbal) e Parte 7 (Criativa/Verbal) – ou noutras com elevadas percentagens de respostas omissas, por serem mais demoradas de executar (as mais difíceis) – Partes 2 (Analítica/Quantitativa), 5 (Prática/Quantitativa) e 6 (Prática/Figurativa). Alguns dos esforços de revisão incidiram, por isso mesmo, na tentativa de redução do nível de dificuldade, designadamente na Parte 2, em que tal se mostrava viável por se tratar de um clássico teste de séries numéricas, e nas Partes 5 e 6, de maneira menos aprofundada, por serem provas baseadas em novo racional. Estes esforços foram bem sucedidos, dentro do possível, muito em especial na Parte 2, que passou de uma das metrologicamente mais fracas na primeira versão, para uma das mais satisfatórias, na última versão.
- A análise de Rasch efectuada no Ensaio Sem Tempo Limite deu, de certo modo, uma leitura à primeira vista mais optimista, posto que apenas um item (da Parte 7, Criativa/Verbal) se mostrou desajustado da medida respectiva, e mesmo assim com um índice de desajustamento pouco expressivo. Note-se que na aplicação sem limite de tempo a generalidade dos índices de discriminação melhoraram substancialmente, e os índices de dificuldade da análise de itens clássica mostraram-se mais equilibrados. Verificou-se, no entanto, fraca calibragem dos itens das Partes Verbais ao nível de competência da amostra, por serem

excessivamente fáceis, o que mostra que se manteve a tendência para o desnível quanto ao grau de dificuldade de diferentes partes da prova.

- A inspecção das respostas aos distractores permitiu, nos vários estudos, identificar alguns itens em que a distribuição das repostas pelas alternativas erradas se mostrou desequilibrada, possivelmente em consequência de ambiguidade na formulação das alternativas. Ainda assim, na análise de Rasch verificou-se que, quando o *STAT-R (H)* é aplicado sem tempo limite, as alternativas de resposta funcionam em geral adequadamente
- Entre os resultados metrológicos destacam-se como mais limitados os que dizem respeito à consistência interna: apesar de ter havido uma assinalável melhoria nos índices relativos aos três domínios de processamento e às três áreas de conteúdo, observando-se os mais favoráveis na sua última versão (2004), eles continuam a situar-se em níveis inferiores aos aconselháveis para utilização das medidas em investigação, a maioria abaixo de .70. A este respeito, a Parte 8 destaca-se do conjunto das restantes ao apresentar sistematicamente os índices de consistência interna mais satisfatórios, a par de outros índices metrológicos também mais adequados (dificuldade e discriminação). E as partes Verbais, especialmente a Parte 1, Analítica, e a Parte 7, Criativa, apresentam-se como as de consistência interna mais baixa, o que pode bem ser explicado pela deficiente calibragem dos itens dessas partes ao nível de competência das amostras. Não será surpreendente que os índices de precisão de cada uma das partes não sejam elevados, dado o número de itens reduzido que incluem; mas a precisão dos resultados compósitos poderia esperar-se ser superior, já que se baseia em 15 itens. Como referido a propósito dos vários estudos de precisão do *STAT-R (H)* (CAPÍTULO 7), Sternberg e colaboradores (Sternberg & Clinkenbeard, 1995; Sternberg, Ferrari, Clinkenbeard, & Grigorenko,1996) sugerem que não seriam de esperar índices muito elevados, em face da heterogeneidade de conteúdo, quer dos domínios (que envolvem sempre três áreas de conteúdo) quer das áreas de conteúdo (que envolvem sempre três domínios de processamento). Com efeito, o princípio da clássica Teoria da Medida de acordo com o qual um elevado nível de consistência interna é imprescindível a qualquer medida psicológica, tem sido posto em causa nas últimas décadas: por exemplo, Raven (1989) ao desafiar uma lista de pressupostos tradicionalmente considerados imprescindíveis em psicometria, refere-se à exigência de elevada consistência interna ou "pureza

factorial" (pp.74-75) como um princípio que deve ser ponderado em cada investigação, à luz da natureza do construto teórico sob avaliação, uma posição assumida também por Messick (2000, p.3): "O nível apropriado de precisão depende do significado do construto que está a ser medido, pois alguns construtos são concebidos teoricamente como mais consistentes ou estáveis do que outros. Assim, a congruência entre a precisão e o significado do construto constitui também uma evidência de validade de construto." Será de acrescentar que, no caso do *STAT-R (H),* este princípio seria aplicável às três áreas de conteúdo, mais do que aos três domínios de processamento, já que estes são considerados conceptualmente mais distintos entre si (contribuindo para a baixa consistência interna das áreas de conteúdo) do que aquelas, o que não se confirma se atendermos ao mais elevado coeficiente que se registou, nos estudos da última versão, na área Quantitativa em comparação com o baixo coeficiente, no domínio Prático, por exemplo.

- Quanto à estrutura interna das medidas importa notar que a tendência global dos resultados foi no sentido de não confirmar o modelo subjacente à construção do *STAT-R (H),* o que desde logo não legitima a interpretação das medidas organizadas em função dos domínios de processamento. A estrutura emergente tendeu a não ser completamente consistente nos diversos estudos mas aproximou-se em geral mais da tradicional organização das aptidões: por um lado, na análise em componentes principais, tende a identificar-se uma componente de natureza geral, com saturações positivas de todas as partes, as mais baixas das quais observadas nas variáveis com mais baixos níveis de consistência interna (e, por inerência, com menores correlações com outras variáveis); por outro lado, quando efectuada a rotação ortogonal dos eixos, os factores organizam-se em torno das áreas de conteúdo mais do que dos domínios de processamento, e dão destaque à área Figurativa, com a qual tende a agrupar-se também a Parte 8, Criativa/Quantitativa, e à área Quantitativa, definida pelas outras duas partes Quantitativas, 2 e 5, Analítica e Prática, respectivamente; a área Verbal, como se poderia esperar, dado o mais baixo nível de precisão, tende a ver diluída a sua importância na explicação da variância dos resultados; por fim, na análise confirmatória o modelo mais ajustado não foi nunca o mais directamente baseado na Teoria Triárquica, mas um modelo hierárquico, baseado em áreas de conteúdo intercorrelacionadas.

- O estudo da dimensionalidade, na amostra em que o *STAT-R (H)* foi aplicado sem limite de tempo e também na do Estudo Principal, demonstrou a presença de importantes fontes de ruído nos resultados, que afectam claramente a sua unidimensionalidade: a variância explicada pelo modelo é, nos vários resultados, quer dos domínios de processamento, quer das áreas de conteúdo, sempre inferior à variância não explicada, e, por consequência, também inferior ao valor critério a partir do qual se admite a possibilidade de unidimensionalidade. Todos os dados apontam, no entanto, no sentido de não ser detectada estrutura nos resíduos, o que constitui evidência de aleatoriedade, isto é, de fontes aleatórias de variabilidade das quais decorre também o baixo nível de precisão das medidas. Assim, apesar do aparente ajustamento dos itens (e também de elevada percentagem de sujeitos) ao modelo de Rasch, o modelo explica apenas uma parte diminuta da variância dos resultados, posto que estes são fortemente afectados por fontes de erro que introduzem ruído nos dados. Daí, a pouca nitidez da estrutura interna das medidas e o grau muito limitado em que essa estrutura se mostra coerente com o modelo teórico que as medidas pretendem operacionalizar.
- A este propósito mostra-se também relevante incluir nesta síntese a referência ao significado metrológico do estudo conjunto das duas técnicas diferenciais. Três constatações efectuadas nesse estudo colocam francamente em questão a validade de construto das medidas desta versão do *STAT-R (H)*: primeiro, a da presença de um factor em que saturam conjuntamente variáveis da *WAIS-III* e do *STAT-R (H)*; segundo, a de que não são as partes Analíticas as que saturam mais claramente nesse factor de natureza geral, mas sim as partes Criativas; terceiro, a de que os resultados da análise confirmatória foram favoráveis à HIPÓTESE EXPERIMENTAL 4.2, a que representa a mais consensual e tradicional estrutura hierárquica das aptidões humanas, que Sternberg, no entanto, postula corresponder a uma gama de funcionamento cognitivo muito limitada, ou pelo menos mais limitada do que a que pretende que seja abrangida pelo *STAT-R (H)*.

O estudo do *STAT-R (H)* deixou claro que, pelo menos na sua versão experimental portuguesa, esta técnica de avaliação da inteligência funcional apresenta limitações metrológicas sérias, o que justifica uma cuidadosa discussão destes resultados.

Em primeiro lugar, pode-se admitir que essas limitações decorreram de deficiente tradução e adaptação; esta explicação seria mais pertinente, contudo, se os resultados provenientes do estudo metrológico da versão original, e também de outras traduções e adaptações internacionais, fossem

francamente superiores aos da versão portuguesa, algo que, como oportunamente se assinalou (ver estudos metrológicos do *STAT-R (H)*, CAPÍTULO 4) não se confirma, nos poucos dados sobre o *STAT* disponíveis na literatura (Sternberg et al. 1996, 1999, 2000, 2001, 2006; ver também Afonso, 2005c). Por outro lado, foram efectuadas sucessivas revisões tendo em vista o aperfeiçoamento da versão portuguesa, sempre com base em estudos empíricos, o que conduziu a uma melhoria sensível, embora a vários títulos insuficiente, da qualidade metrológica das medidas, ao contrário do que se passou com a *WAIS-III* – cuja tradução foi efectuada no âmbito deste mesmo projecto e conduziu, desde a primeira versão experimental, a resultados metrológicos bastante satisfatórios.

Em segundo lugar, não deixa de ser oportuno manifestar alguma perplexidade relativamente à reduzida publicação de resultados suficientemente completos e claros de estudos metrológicos do *STAT-R (H)*, ou seja, da sua última revisão (2002), apesar de continuar a ser citado a propósito da validação interna da Teoria Triáquica (e até mesmo do modelo *WICS*) (ver, por exemplo, Sternberg, 2005d) e de constituir uma das técnicas utilizadas em investigação, no âmbito do *PACE Center* (*Psychology of Abilities, Competencies and Expertise Center*)[5], muito designadamente no projecto *Rainbow*, onde tem sido aplicado, juntamente com novos métodos de avaliação (testes de desempenho), e em complemento das técnicas tradicionais de selecção de candidatos ao ensino superior americano (do *SAT- Scholastic Assessment Test; o ACT – American College Test; o GRE – Graduate Record Examination; e o GMAT – Graduate Management Admissions Test*) (Stemler, Grigorenko, Jarvin & Sternberg, 2006; Sternberg & The Rainbow Project Collab., 2006; Sternberg, The Rainbow Project Collab. & the University of Michigan Business School Project Collab., 2004).

Alguma fragilidade no rigor e na consistência da investigação empírica efectuada e coordenada por Sternberg, que para alguns contrasta com a extraordinária criatividade e produtividade da sua obra teórica[6], tem sido alvo, aliás, de criticas que é oportuno ponderar, a propósito dos resultados obtidos nesta investigação. Brody (2003a,b) aponta as seguintes limitações aos estudos de validação da Teoria Triárquica com recurso ao *STAT (H)* (1993) (Sternberg, Castejón, Prieto, Hautamäki & Grigorenko, 2001; Sternberg, Ferrari, Clinkenbeard & Grigorenko, 1996; Sternberg, Grigorenko, Ferrari & Clinkenbeard, 1999):

- As correlações alegadamente baixas entre os domínios de processamento do *STAT (H)* e testes de inteligência mais ou menos convencionais (Sternberg et al.,1996) terão sido afectadas por dois factores, o baixo nível de precisão das medidas do *STAT-R* (que tem um efeito de atenuação) e a restrição da amplitude

[5] O centro de investigação fundado e dirigido por Sternberg e actualmente localizado na Universidade de Tufts (Massachussets, USA).

[6] "As conclusões do Professor Sternberg seriam mais aceitáveis para os investigadores do domínio da inteligência se o rigor das suas análises de dados igualasse a originalidade dos seus *insights* teóricos." (Brody, 2003b, p.342).

dos resultados (baixa variabilidade devida a homogeneidade das amostras de estudantes, seleccionados com base em critérios de sobredotação). De facto, nos estudos de Sternberg e colaboradores, as correlações entre variáveis não são habitualmente corrigidas quanto à atenuação (devida a baixa precisão), embora sejam discutidas quanto à sua magnitude e comparadas entre si; de acordo com Sternberg (2003f), a aplicação de correcções baseia-se no pressuposto de que a correlação entre duas variáveis é a mesma ao longo de toda a amplitude de resultados, o que é sabido não ser exacto (as correlações entre testes de aptidão, por exemplo, são mais elevadas nos níveis mais baixos de competência do que nos níveis mais elevados). A correcção da restrição de amplitude, por seu lado, requer a correcção para a "população correcta", o que nem sempre é delimitável com exactidão, e corrigir para uma população inapropriada é arbitrário e não reflecte a situação ecológica para a qual foi criado o teste (Sternberg, 2003f, p.332); o autor prefere, por isso, trabalhar com amplitudes "reais", de amostras de estudantes "reais", e não com amplitudes artificialmente inflacionadas, relativas a hipotéticas amostras que contêm toda a gama de competências concepualizável (Sternberg & Williams, 1998, p.576; ver também Sternberg & The Rainbow Project Collab, 2006).

- O domínio Criativo apresenta uma correlação ligeiramente mais elevada do que o domínio Analítico com os testes convencionais de inteligência, que de acordo com a Teoria Triárquica medem a inteligência analítica; ao referir-se às correlações com outros testes de inteligência, Sternberg (2003f) sublinha que o padrão de correlações, mesmo não sendo perfeito, se configura em geral de acordo com o previsto e que as mais elevadas correlações do domínio Criativo com testes de factor g (como o *Cattell Culture-Fair Test*) se devem a que estes, apesar de medirem competências analíticas, o fazem de "maneira nova". Em observação de tom irónico, Brody (2003b, p.340) afirma que este argumento denuncia uma mudança assinalável na linha de argumentação de Sternberg: as medidas convencionais de *g*, encaradas na Teoria Triáquica como medidas da inteligência analítica, não da inteligência criativa nem da inteligência prática, passam a ser tidas, enquanto medidas da inteligência fluida, como melhores medidas da inteligência criativa do que da inteligência analítica. Julga-se oportuno acrescentar que, nesta perspectiva, o contraste entre inteligência analítica e inteligência criativa, a observar-se, pode eventualmente ser lido como correspondendo à

distinção entre inteligência cristalizada e inteligência fluida, há muito introduzida por Cattell e por Horn e mais recentemente incorporada por Carroll na Teoria dos Três Estratos (ver CAPÍTULO 1, p.47), distinção que é, afinal, coerente com a que repetidamente é feita por Sternberg entre as inteligências analítica e criativa, a partir do carácter familiar, e próximo das aprendizagens escolares, das tarefas envolvidas na primeira e do carácter não familiar (logo, desligado das aprendizagens escolares) das tarefas envolvidas na segunda[7].

- O padrão de relações entre os resultados do *STAT (H)* e o desempenho académico não é tão nítido quanto pretendem os autores (Sternberg et al., 1999): as correlações entre os três domínios do *STAT (H)* e as avaliações académicas respectivas – analítica, prática e criativa – não são significativamente mais elevadas do que as correlações cruzadas, entre resultados do *STAT(H)* e avaliações académicas com designações diferentes; além disso, o resultado Analítico explica sozinho mais de 75% da variância explicada pelo conjunto dos três domínios, o que significa, por outras palavras, que a predição do rendimento académico nos diferentes domínios, prático e criativo incluídos, é apenas marginalmente melhorada pela consideração da combinação dos três resultados do *STAT (H),* em substituição do resultado Analítico isolado. Embora Sternberg (2003f) argumente que esta objecção repousa numa confusão de variáveis dependentes *entre* condições de tratamento com variáveis dependentes *dentro* das condições de tratamento, Brody (2003b) sublinha que as medidas do *STAT (H)* não fornecem suficiente predição das avaliações académicas, pondo em causa a sua validade de construto, ao que acresce não terem sido comparadas com predições efectuadas a partir de medidas convencionais de inteligência (do ponto de vista da validade incremental).
- As correlações significativas entre os três domínios de processamento do *STAT(H)* não são indicativas de independência relativa, mais ainda se a atenuação devida aos baixos índices de precisão e a restrição da amplitude forem corrigidas. Sternberg atribui essas correlações a variância de método – variância comum devida às competências envolvidas na resolução de testes de resposta múltipla – pelo que remove o seu efeito, de onde resultam correlações próximas de zero. A

[7] Comente-se, a propósito, que, embora Sternberg repetidamente afirme que as concepções tradicionais de inteligência excluem a "inteligência criativa", definida na subteoria experiencial como capacidade para lidar com a novidade, há que assinalar que não só o conceito de inteligência fluida (Gf) introduzido por Cattell envolve a aptidão para resolver problemas novos como ainda, tanto o modelo de Cattell como o de Carroll incluem um factor de Fluência Geral (Gr) cuja definição contempla também a noção de criatividade.

mesma covariância pode ser igualmente interpretada como representando g (Brody, 2003a,b), e nada terá de surpreendente a constatação de que ao remover das correlações a principal fonte de covariação (eventualmente *g*) elas se aproximam de zero; para Sternberg, a interpretação desta fonte de covariação, como variância de método, ou como variância de construto, é meramente uma questão de opinião, e a sua opinião é, sem dúvida, a de que se trata de variância de método (Sternberg, 2003f, 2006b). Convém acrescentar que, apesar disso, Sternberg (2006b; Sternberg & The Rainbow Project Collab., 2006) tem vindo a admitir mais explicitamente a proximidade entre as medidas do *STAT-R (H)* e as técnicas tradicionais de avaliação da inteligência, sobretudo, medidas da inteligência fluida.

- A hipótese de que os estudantes expostos a formas de instrução compatíveis com o seu modo dominante de processamento mental – analítico, prático ou criativo – têm mais sucesso do que os estudantes que não usufruem dessa compatibilidade (Sternberg at al., 1999), também não ficou demonstrada, de acordo com Brody (2003a), uma vez que cerca de 35% da amostra foi excluída da análise, sob pretexto de eliminar as flutuações aleatórias decorrentes do impacto de factores não académicos nos resultados, sendo que essa exclusão de participantes com "resultados desviantes" (Sternberg et al., 1999, p.10) obedeceu a critérios não explícitados. Este procedimento seguramente reduziu a variância *dentro* dos grupos, embora os dados não permitam determinar se afectou também a variância *entre* os grupos; mas poderá ter contribuído para a elevação das estatísticas de teste (F).
- Por fim, os resultados da análise factorial confirmatória (Sternberg et al., 2001) não são suficientemente coerentes com a teoria, uma vez que a superioridade do ajustamento das medidas ao modelo Triárquico foi obtida à custa de permitir que os factores de segunda ordem, Analítico, Prático e Criativo, se correlacionassem (média das correlações = .855); esta opção corresponde a admitir a presença de um factor comum aos três domínios de processamento, que não só não se mostram, portanto, independentes entre si, como previa a Teoria Triárquica, como podem ainda conceber-se dependentes de um único factor geral. Acresce que a análise factorial confirmatória do *STAT (H)* isolado não constitui, de acordo com Brody, o melhor método para demonstrar a ausência de variância de g nos resultados: o estudo correlacional conjunto do *STAT (H)* e outras medidas

convencionais de inteligência constitui um meio mais adequado à identificação do lugar das medidas do *STAT (H)* na estrutura das aptidões humanas, sugestão que, recorde-se, inspirou a presente investigação (ver definição do problema no CAPÍTULO 4).

- Em suma, Brody sublinha que a presença de variância de *g* nos resultados do *STAT (H)* é apoiada por três ordens de evidências empíricas: a correlação entre cada domínio de processamento e medidas convencionais de *g*, as correlações substanciais entre os três domínios, e os resultados da análise factorial confirmatória.

É importante notar que nos estudos a que se reporta Brody na sua crítica foi utilizado o Nível H do *STAT,* a forma mais antiga do teste, em utilização desde 1993 (uma versão com 36 itens de resposta múltipla, 4 em cada parte, e 3 itens "ensaio", um para cada domínio). Não são, por isso, directamente comparáveis os resultados dessa versão do teste e os resultados da presente investigação; ainda assim, e porque são esparsos os resultados metrológicos da versão revista americana do *STAT-R (H)* (2002) (Sternberg & The Rainbow Project Collab., 2006; Sternberg, The Rainbow Project Collab. & the University of Michigan Business School Project Collab., 2004), vale a pena assinalar algum paralelismo entre as considerações de Brody (2003a,b) e os mais importantes resultados obtidos com a versão portuguesa do *STAT-R (H)* (2004): 1) correlações apreciáveis (.50 a .71) entre os domínios de processamento e o resultado compósito (Escala Completa) da *WAIS-III*, uma medida tradicional da "inteligência geral" (bem como entre as partes do *STAT-R (H)* e o subteste de Matrizes – sete das nove correlações significativas, .28 a .57); 2) em particular, maiores correlações entre as partes Criativas, do que entre as partes Analíticas ou Práticas, e os resultados da *WAIS-III*; 3) correlações expressivas, mesmo sem correcção de atenuação ou de restrição de amplitude, entre os três domínios de processamento; e 4) como consequência do precedente, melhor ajustamento, na análise confirmatória, a modelos em que os domínios não são independentes, os quais permitem supor a presença de um factor comum aos três domínios de processamento.

A emergência de uma componente geral, passível de ser interpretada como variância devida a *g*, na estrutura interna do *STAT*, reveste-se da maior importância, do ponto de vista conceptual, como do ponto de vista da validação das medidas da inteligência funcional. Conceptualmente, esta constatação faz cair pela base um dos mais fundamentais postulados da Teoria Triárquica, o de que "*g* não é geral", apenas *parece* geral porque é partilhado pelo conjunto limitado de tarefas e problemas de natureza "académica" tradicionalmente incluídos na avaliação da inteligência, mas pouco relevantes para o sucesso na vida comum, já que esta exige outras competências como a inteligência prática e a inteligência criativa. As correlações elevadas entre os três domínios e o facto de emergir, de maneira

consistente, em diferentes amostras, quer com a forma portuguesa quer mesmo com a forma original do *STAT-R (H),* um factor comum, permite, pelo contrário, admitir que meramente se está perante novas formas de medir uma mesma inteligência, numa acepção bem próxima de Wechsler, ou quanto muito perante duas formas de inteligência correlacionadas, inteligência cristalizada e inteligência fluida, na acepção de Cattell e Horn, não perante novas formas de inteligência independentes, na acepção de Sternberg. Daí que, do ponto de vista da validação das medidas, este tipo de resultados conduza à rejeição da interpretação originalmente proposta para os resultados proporcionados pelo *STAT-R (H)* – medida de três domínios de processamento correspondentes a três formas de inteligência, Analítica, Prática e Criativa – estabelecendo-se antes como mais legítima a interpretação do resultado compósito obtido a partir do conjunto de todos os itens, como uma medida da inteligência geral.

Não havendo correspondência entre o modelo teórico e a sua operacionalização, duas leituras se tornam possíveis: poderá ser que o *STAT-R (H)* não operacionalize adequadamente a Teoria Triárquica da Inteligência, ou poderá ser que a própria Teoria Triárquica da Inteligência não encontre suficiente suporte empírico, pelo menos a partir da investigação com o *STAT*? Enquanto esta última perspectiva tende a caracterizar a postura dos críticos a Sternberg, desde os mais radicais, que contestam a teoria na sua globalidade (Gottfredson, 2003a,b) aos mais moderados, que apenas a contestam parcialmente (Brody, 2003a,b) ou consideram que não se demarca suficientemente das concepções tradicionais de inteligência (Gardner, 1983), o autor e seus colaboradores tendem a manter que os resultados, embora longe de serem perfeitos, se mostram promissores quanto ao estabelecimento de uma nova forma de conceptualizar e avaliar a inteligência (Sternberg, 2003f,h), razão por que persistem na demonstração da sua utilidade para a avaliação da inteligência baseada na Teoria Triárquica (Sternberg et al., 2004, 2006).

Reserva-se para o tópico 9.2 do presente Capítulo, a abordagem das implicações dos resultados desta investigação do ponto de vista da validação das concepções teóricas estudadas e enquadradas nos paradigmas diferencial e sistémico de investigação da inteligência; para completar a reflexão sobre as implicações dos resultados para a medição e investigação da inteligência, importa por agora analisar as evoluções recentes das perspectivas de medida da inteligência, tomando por referência os paradigmas diferencial e sistémico de investigação, e a posição que ocupam actualmente, e que poderão vir a ocupar, as duas técnicas de avaliação da inteligência estudadas, no quadro da evolução da medida e da investigação da inteligência que se antecipa para as próximas duas a três décadas.

9.1.3. *WAIS-III* e *STAT-R (H)* no Quadro da Evolução da Medida da Inteligência

Com alguma frequência, Sternberg manifesta a sua perplexidade perante a estagnação da tecnologia de avaliação psicológica, sobretudo ao tomar como termo de comparação a evolução de outras tecnologias ao longo do século XX:

> o "Nenhuma tecnologia de que eu tenha conhecimento – computadores, telecomunicações, televisões, e outras – terá apresentado o tipo de estagnação de ideias que tem caracterizado a indústria da medida psicológica. Porquê? Porque noutras indústrias, quem não inova, não sobrevive. Na indústria da medida psicológica, o oposto parece ser verdade. Como nos filmes *Rocky I, Rocky II, Rocky III,* etc., a indústria da medida psicológica vai sucessivamente desenvolvendo variantes cosméticas menores do mesmo produto, em que a única mudança substancial consiste no número que sucede o nome do teste. Estas variantes sobrevivem porque os psicólogos compram os testes e de seguida os defendem lealmente […]. Os testes existentes e as suas utilizações comuns têm valor, mas não constituem o que de melhor poderia fazer-se. Quando um comentador afirma que «não será jamais possível melhorar substancialmente o essencial das políticas de admissão de candidatos [ao ensino universitário]» (Darlington, 1998, p.572), a afirmação é equivalente à que alguns terão proferido a propósito do automóvel *Modelo T* ou do computador *UNIVAC*" (Sternberg & Williams, 1998, p.577).

Com efeito, se nos detivermos na consideração das grandes linhas de evolução de técnicas diferenciais clássicas como as baterias de Wechsler, um trabalho que foi efectuado relativamente à *WAIS* com base na análise dos manuais das três edições, 1955, 1981 e 1997 (Afonso, 2004a), verificamos que a evolução dos procedimentos metodológicos e nas técnicas de análise de resultados empregues no estudo metrológico, evidentemente apoiada pelos desenvolvimentos da estatística, pela aplicação de modelos matemáticos à investigação psicológica e pelo recurso às tecnologias informáticas de tratamento de dados, não tem paralelo no plano conceptual, onde persiste, na essência, a concepção de inteligência global de Wechsler, com cerca de setenta anos, ela própria enraizada em perspectivas e conceitos com mais de um século (Binet & Henri, 1895; Spearman, 1904).

Esta fidelidade às ideias originais de Wechsler merece alguma reflexão. Em primeiro lugar, sem dúvida sublinha a actualidade das noções introduzidas pelo autor em 1939, na primeira edição da sua obra *The measurement and appraisal of Adult Intelligence,* e por ocasião da publicação da sua primeira escala de inteligência, a *Wechsler-Bellevue I*: noções como a de que a inteligência se desenvolve para lá do final da adolescência, a de que a inteligência depende do funcionamento global

do cérebro não tendo localização específica, a de que a inteligência consiste numa manifestação da personalidade como todo, não sendo imune, consequentemente, à influência de factores "conativos" ou "não intelectivos", são de uma extrema actualidade se considerarmos até, especificamente, algumas das linhas de evolução actual da investigação da inteligência humana – as relações entre aspectos cognitivos e conativos do funcionamento psicológico, ou a investigação neuropsicológica, em particular o estudo dos correlatos neurofisiológicos do funcionamento cognitivo, por exemplo, constituem tópicos de extrema actualidade na Psicologia. Mas também é verdade que a constatação desta fidelidade às ideias de Wechsler não pode perder de vista o impacto comercial e a enorme divulgação das escalas deste autor (as mais utilizadas técnicas de avaliação psicológica a nível internacional) as quais mantêm deste modo uma identidade muito própria no domínio da medida da inteligência, em diversos contextos de intervenção do psicólogo (Afonso, 2004a). Acresce ainda que a este sucesso não será também alheio o enorme corpo de investigação que apoia a noção de inteligência como construto global, indiscutivelmente um dos maiores sucessos da Psicologia Científica (Gottfredson, 2003a; Nyborg, 2003), o qual dá enquadramento, também, aos mais importantes resultados obtidos na presente investigação, muito designadamente com a *WAIS-III* em amostras portuguesas. Ainda assim, não deixa de ser surpreendente, como Sternberg faz notar, que a tecnologia da medida psicológica não tenha acompanhado a evolução que marcou o século XX, nos mais diversos domínios.

Apesar da critica velada de Sternberg às técnicas de avaliação da inteligência do tipo das escalas de Wechsler (como a *WAIS-III*), paradoxalmente, Sternberg oferece como técnica de avaliação da inteligência construída com base no seu modelo sistémico um instrumento que partilha muitas características das técnicas mais clássicas de avaliação colectiva da inteligência: trata-se de uma técnica de papel e lápis e de aplicação estandardizada (ou seja, que procura controlar, ou mesmo eliminar, o efeito do contexto nos resultados), de resposta por escolha múltipla (ou seja, que impede a manifestação da idiosincrasia individual no comportamento de resposta e inviabiliza qualquer apreciação qualitativa do processamento envolvido na resolução da tarefa), de desempenho máximo (ou seja, que assume a classificação dicotómica das respostas em "certas" e "erradas", de onde decorre que as diferenças individuais se expressem em termos quantitativos de "grau" ou de "nível" de aptidão em cada dimensão medida, isto é, em cada domínio de processamento), centrada nos produtos (resultados quantitativos) mais do que nos processos (descrição qualitativa do processamento mental), envolvendo conteúdos clássicos (verbal, quantitativo e figurativo) e alguns tipos de tarefas também clássicas (matrizes, séries, analogias verbais, problemas aritméticos, etc.) e, acima de tudo, que se centra na avaliação de factores do mundo interno do indivíduo, mesmo nas partes do teste supostamente contextuais (Partes Práticas) em que, em parte como consequência da opção por fazer representar apenas o processo de adaptação, e não os de modelação e de selecção do meio, a ênfase

é colocada na "capacidade" do indivíduo para se modificar no sentido de se conformar ao contexto, como dado adquirido (a noção clássica, afinal, de inteligência como adaptação ao meio em sentido lato, não ao meio imediato do indivíduo). Esta última característica, acrescentada ao formato de escolha múltipla, implica a perda da possibilidade de avaliação do aspecto mais essencial da inteligência prática, a flexibilidade da resposta em função das circunstâncias, aspecto nuclear, afinal, da subteoria contextual. O formato de resposta por escolha múltipla é, aliás, também pouco compatível com a avaliação da inteligência criativa: neste caso, perde-se a possibilidade de uma verdadeira avaliação do pensamento criativo, por definição inovador e imprevisível, ao formatar as respostas como alternativas *a priori,* e classificáveis em "certas" e "erradas"[8]; acrescente-se que a inclusão no *STAT* apenas de um dos segmentos do contínuo novidade/automatização, de problemas que supostamente requerem lidar com a novidade, mas não de problemas que requerem automatização, contribui para reduzir a validade de construto das medidas derivadas da subteoria experiencial (Partes Criativas): decorre daqui que está em causa, nestas partes do *STAT*, avaliar a "capacidade" para lidar com a novidade, de novo um conceito dimensional que remete para o mundo interno do indivíduo e que é avaliado por um resultado expresso na forma quantitativa.

É justo que se sublinhe que estes comentários dizem respeito às nove partes do *STAT-R (H)* de resposta por escolha múltipla, mas que o autor, sensível a questões como as que acabam de ser tratadas, desde o início procurou introduzir também medidas mais qualitativas, designadamente na Parte 10 da prova original, constituída por três itens-ensaio[9]. Nesta parte, cada item requer a redacção de um texto ou ensaio, de carácter analítico (análise das vantagens e desvantagens do policiamento das escolas), de carácter prático (proposta de três soluções práticas para um problema da vida pessoal actual) ou de carácter criativo (descrição da "escola ideal") e a cotação é efectuada por juízes treinados que classificam as respostas em função de um conjunto de critérios. Ao introduzir itens-ensaio no *STAT (H)*, Sternberg estava ciente das limitações do formato de resposta por escolha múltipla, sobretudo da limitação decorrente da variância de método (como referido antes, o autor considera que o simples formato de resposta por escolha múltipla envolve o recurso a um conjunto de competências relativamente independentes dos construtos que se pretende medir, mas que se reflectem nas intercorrelações das medidas); sublinha por isso, que as correlações entre os itens de resposta por escolha múltipla são bem mais elevadas do que as correlações entre estes e os itens ensaio (ver, por exemplo, Sternberg et al, 1999; Sternberg, 2005e), o que é interpretado por alguns autores (Brody,

[8] Essa pode ser uma das razões por que as Partes Criativas tendem a apresentar elevadas correlações com as Partes Analíticas e com os testes tradicionais de inteligência.

[9] Como referido e justificado oportunamente (cf CAPÍTULO 4, p..281) na presente investigação optou-se por não utilizar a Parte 10.

2003a,b) como resultado do baixo nível de precisão dos itens ensaio (apenas um item para avaliar cada domínio de processamento).

Apesar de ciente das limitações do formato de resposta por escolha múltipla, a verdade é que Sternberg construiu o *STAT,* que ainda hoje continua a utilizar no âmbito do *Rainbow Project*, predominantemente com esse formato (45 dos 48 itens são de escolha múltipla). Que vantagens poderia antecipar nessa opção? Será sem duvida especulativo discernir os motivos de Sternberg a este respeito, uma vez que ele próprio nunca os apresentou de maneira explícita. Mas pode-se admitir que procurasse reter as vantagens conhecidas desse formato – a facilidade de aplicação e a economia de tempo e de recursos decorrente de aplicação colectiva, a possibilidade de classificação automática e a inerente objectividade de cotação, e o controlo de fontes de erro contextuais através da aplicação estandardizada – às quais se juntassem, talvez, motivos de ordem comercial, já que houve desde cedo interesses editoriais na publicação do *STAT (H)* e um teste de formato convencional teria maior probabilidade de alcançar a aceitação por parte da empresa editora e a popularidade entre os psicólogos. Convém também não esquecer que a primeira versão do *STAT* foi concebida para utilização no âmbito de um estudo específico, o primeiro em grande escala conduzido com o propósito de validação da Teoria Triáquica, no qual havia que avaliar, a nível nacional, alunos do ensino secundário no âmbito de um processo de avaliação dos estudantes que iriam frequentar um curso de Verão, na Universidade de Yale (ver, por exemplo, Sternberg & Clinkenbeard, 1995; Sternberg, Ferrari, Clinkenbeard, & Grigorenko, 1996). Compreende-se, assim, a opção por uma prova de aplicação colectiva e de resposta por escolha múltipla. Mas as limitações metrológicas detectadas, quer na versão original quer em versões internacionais, a ambiguidade quanto à natureza da variância comum aos itens de escolha múltipla (variância de método *versus* variância de construto) e considerações acerca da natureza dos construtos que incorporam a Teoria Triárquica terão determinado o crescente empenho de Sternberg e colaboradores no desenvolvimento de novos meios de avaliação da inteligência funcional, o que deu origem, nos últimos seis anos, a novos testes especialmente concebidos para a avaliação das inteligências prática e criativa (Grigorenko & Sternberg, 2001; Sternberg, 2005d; Sternberg & The Rainbow Project Collab., 2006 ; Sternberg, The Rainbow Project Collab., & The Univ. of Michigan B.S.P. Collab., 2004):

- <u>novas medidas da inteligência prática</u>:

  1) Técnicas de auto-descrição compreendendo itens que pretendem avaliar as competências práticas no domínio social (por exemplo, utilização eficaz da comunicação interpessoal), no domínio familiar (por exemplo, como consertar objectos domésticos ou como gerir o orçamento familiar) e no domínio da resolução de problemas inesperados (por exemplo, como organizar algo que se tornou caótico);

2) Vinhetas (gravuras) acompanhadas de cinco alternativas das quais há que escolher a que se considera melhor para resolver quatro tipos de problemas: como manter o valor das poupanças, o que fazer quando após uma compra se verifica que o objecto comprado está avariado, como conseguir assistência médica em época de crise ou como gerir um bónus salarial que se recebeu pela excelência do trabalho efectuado. Não existe resposta "certa" e a resposta "chave" foi determinada a partir da frequência de resposta (a mais frequente);

3) Inventário de Avaliação Situacional: consiste em sete situações apresentadas na forma de *video* as quais lidam com problemas do quotidiano, como o que fazer se alguém que não conhecemos muito bem nos pede para escrever uma carta de recomendação;

4) Questionário de Senso Comum: composto por 15 vinhetas representando problemas que podem ser enfrentados em contextos de trabalho, como lidar com uma tarefa fastidiosa ou resolver uma situação competitiva;

5) Questionário da Vivência Universitária: composto por 15 vinhetas que representam problemas que podem ser enfrentados no contexto universitário, como providenciar a obtenção de uma bolsa ou lidar com um companheiro de quarto difícil.

- novas medidas da inteligência criativa:

1) Descrever o mundo através dos olhos dos insectos;

2) Descrever quem vive e o que acontece num planeta chamado *Priumliava*, sem qualquer informação adicional;

a cotação de qualquer destas duas tarefas é feita por juízes teinados e incide na originalidade (inovação), no desenvolvimento ideativo (qualidade) e na utilização de conhecimento prévio numa tarefa relativamente nova (sofisticação);

3) Banda desenhada: cinco imagens de banda desenhada das quais o sujeito deve escolher três para preencher as falas das personagens;

a cotação é feita por três juízes treinados e incide na sagacidade, no humor e na originalidade;

4) Histórias escritas: redigir duas histórias, demorando cerca de 15 minutos com cada uma, escolhendo dois dos seguintes títulos: "A quinta oportunidade", "2983", "Para lá do precipício", "Os chinelos do Polvo", "Está a recuar" e "Sem tempo suficiente";

a cotação é feita por quatro juízes treinados e incide na originalidade, na complexidade, no poder de evocação da emoção e no poder descritivo;

5) Histórias Orais: perante a apresentação de cinco folhas contendo gravuras relacionadas com um determinado tema comum – música, dinheiro, viagem – o sujeito tem de escolher duas e em 15 minutos, para cada tema, formular uma história e ditá-la, no máximo de 5 minutos, para um audio-gravador;

a cotação é feita por seis juízes treinados e incide nos mesmos critérios do teste anterior.

De acordo com Sternberg (2005d) este tipo de testes não só faz emergir claramente os factores prático e criativo separados de factor analítico, medido por técnicas convencionais de avaliação da inteligência, como se mostra preditivo do sucesso em contextos práticos da vida (Grigorenko & Sternberg, 2001) ou na adaptação à universidade (Stemler, Grigorenko, Jarvin & Sternberg, 2006; Sternberg, 2005d; Sternberg & The Rainbow Project Collab., 2006). Estes novos formatos de avaliação contrastam com os testes de resposta por escolha múltipla, desde logo por serem mais contextuais, colocando tanto quanto possível o sujeito perante situações que poderia enfrentar no seu quotidiano; por serem de resposta aberta, o que torna a avaliação mais flexível, mais passível de revelar a idiosincrasia individual e de fornecer informações qualitativas quanto aos modos de resolução das tarefas; por serem de comportamento típico, não de desempenho máximo - as que se baseiam na auto-descrição e evitam a imposição de uma classificação rígida em respostas certas e erradas; por serem mais preocupadas com a activação dos processos na situação de avaliação do que com a medição dos produtos. Ainda assim, persiste a concentração da avaliação da inteligência no mundo interno do indivíduo (as situações ou contextos constituem pretextos para se observar o desempenho *do indivíduo*, ou para avaliar o funcionamento *do indivíduo,* não actuando este sobre o contexto durante a resolução do problema) e são semelhantes para todos os indivíduos; além disso, mantém-se o peso da utilização da linguagem, da qual depende em alguma parte a qualidade das respostas, sendo excessiva a presença de tarefas de redacção de textos[10]. A maior desvantagem destas técnicas reside, contudo, na morosidade da sua classificação que requer a cuidadosa definição de critérios, o recurso a vários juízes previamente treinados e o controlo cuidadoso da consistência da cotação inter-avaliadores. Será talvez oportuno recordar que os resultados do estudo do acordo inter-avaliadores em testes tradicionais de inteligência, como a *WAIS-III*, mostram-se em geral encorajadores e permitem atingir elevado nível e consistência, quando os critérios são estabelecidos com cuidado e aperfeiçoados a partir de estudo empírico. Mas a extensão das respostas é na *WAIS-III* muito mais curta e os critérios envolvidos menos diversificados e complexos.

[10] Convém que se refira que boa parte destas técnicas foram criadas para utilização no âmbito do projecto *Rainbow*, na selecção de candidatos ao ensino superior.

Nestas novas formas de avaliação da inteligência manifestam-se sobretudo três ênfases: a configuração como técnicas de desempenho típico, em contraste com as de desempenho máximo, o abandono do formato de resposta por escolha múltipla, substituindo-o pelo de resposta aberta, e a tentativa de maior aproximação da situação de observação a contextos representativos da vida quotidiana. Quanto ao primeiro aspecto, importa notar que as medidas assumem carácter mais idiográfico, centrado na compreensão do funcionamento individual, do que nomotético, focado na quantificação do produto desse funcionamento, e implicam, por consequência, atenção à avaliação subjectiva do sucesso a par da medição objectiva do sucesso. Ao reportar as medidas desta natureza à concepção de inteligência que lhes deu origem, faz sentido que assim seja: definir a inteligência como "capacidade do indivíduo para, de forma intencional, adaptar-se ao meio, modelar o meio ou seleccionar um meio, tendo em vista satisfazer os seus objectivos pessoais e os objectivos que a sociedade ou a cultura em que vive lhe impõem" faz sobressair que "os traços associados com a inteligência funcional são parcialmente idiográficos, não nomotéticos" (Sternberg, 2001b, p.350) e que medir a inteligência tende a ser um exercício de avaliação global do ajustamento, o qual inclui também a vivência subjectiva do ajustamento (Afonso, 2005b).

Este tipo de avaliação, que não substitui, antes complementa, a avaliação tradicional com recurso a técnicas de desempenho máximo (como se viu antes, inspiradas numa visão do mundo mecanicista), pode ser entendido como mera formalização de uma prática desde há muito implantada na avaliação psicológica, a de associar num mesmo processo a avaliação de condutas cognitivas e conativas, no quadro de um entendimento articulado e interdependente de diferentes áreas de funcionamento psicológico. Note-se que, deste ponto de vista, não se trata de um entendimento em si mesmo inovador, como se viu a propósito da concepção de inteligência de Wechsler; mas sim de um entendimento que remete para uma visão do mundo selectivista, que sublinha a intencionalidade do comportamento e reporta a acção aos objectivos e valores pessoais do indivíduo. Mas esta ideia é ainda mais elaborada por alguns autores: a aposta de muita da tradicional avaliação da inteligência na medição encarada como independente dos valores e objectivos pessoais é incluída por Raven (1989) na lista de postulados da psicometria clássica que considera no mínimo questionáveis: "não faz sentido tentar aceder separadamente às componentes cognitivas, afectivas e conativas de uma actividade. Não se pode avaliar, de um modo que faça sentido, a capacidade de desenvolver melhores maneiras de pensar sobre algum problema, independentemente do prazer que a pessoa em causa obtenha ao fazê-lo, ou da sua determinação em tornar conhecidas as suas ideias" (p.74). Raven sugere, por consequência, que a avaliação da competência cognitiva deve ser efectuada em dois estádios: no primeiro são identificados os tipos de comportamento ou actividade que o indivíduo valoriza, e só de seguida se avalia a maneira como aplica e integra as suas potencialidades cognitivas, afectivas e

conativas na implementação eficaz desse tipo de comportamento ou actividade. Raven sublinha que não se trata de medir em separado, como é prática corrente, os aspectos cognitivos, afectivos e conativos para depois os integrar, no momento da interpretação, mas antes de formatar a própria avaliação cognitiva a partir da avaliação prévia das vertentes afectiva e conativa da cognição – o que pressupõe, ao que parece, o desenvolvimento de métodos de "avaliação adaptativa", em que a selecção dos itens (ou das próprias tarefas) a ser aplicados depende não apenas do *nível* de competência previamente demonstrada, o tipo de avaliação adaptativa mais comum, mas do *tipo* de competência valorizada ou investida pelo próprio sujeito. A ênfase excessiva nos objectivos e valores pessoais, da qual poderia decorrer desajuste relativamente às pressões do contexto, é contrabalançada por Raven ao questionar um outro princípio da psicometria clássica, o de que a avaliação pode ser independente do contexto em que se encontra o indivíduo, na dupla acepção do contexto objectivo (que terá de ser também avaliado do ponto de vista das oportunidades de aprendizagem que oferece e dos desafios que coloca) e do contexto percebido ou representação subjectiva do contexto. Resta decidir se toda esta ampliação do conceito e da medida da inteligência não poderá conduzir à diluição dos contornos do construto, relativamente a outros construtos psicológicos, e, por inerência, à redução das possibilidades de implementação de procedimentos rigorosos e consequentes de validação das medidas.

Quanto à segunda ênfase das novas formas de avaliação – opção pelo formato de resposta aberta, para a medida das inteligências prática e criativa – há que reconhecer que se trata também de uma ideia não completamente nova: as mais clássicas baterias de inteligência de aplicação individual, como as escalas de Wechsler, incluem uma maioria de testes de resposta aberta[11] que, apesar de tornarem mais morosa, e eventualmente menos precisa, a classificação, nunca foram abandonados, por proporcionarem preciosa oportunidade de observação do comportamento espontâneo em situação de avaliação. Os itens que constituem as novas medidas construídas por Sternberg e colaboradores parecem exigir, no entanto, um enorme dispêndio em recursos humanos: os textos ou os registos áudio são longos, em comparação com as respostas aos subtestes das provas de Wechsler, e, depreende-se, demorados de apreciar e classificar, e o recurso a vários juízes torna as técnicas pouco adaptadas à utilização para lá do contexto estrito da investigação. Enquanto não forem desenvolvidas metodologias mais rentáveis de classificação das respostas abertas, a validação destas técnicas terá forçosamente de tomar em consideração a relação custo/benefício, no quadro da chamada validade incremental por referência a alternativas de avaliação da inteligência menos exigentes em termos de recursos humanos.

[11] Alguns dos quais poderia até ser invocado medirem aspectos da inteligência funcional, como o subteste de Compreensão, que em parte envolve a avaliação da capacidade de resolução de situações de ordem prática.

Uma terceira característica que estes novos métodos de avaliação da inteligência apresentam, que os distingue da maioria das técnicas tradicionais (principalmente, colectivas), consiste na tentativa de maior aproximação a situações e contextos do quotidiano: pode-se ler esta tendência como um prolongamento de uma mudança iniciada por Binet, ao argumentar a favor da medida da inteligência como construto global, efectuada a partir da eficiência na resolução de tarefas e problemas representativos do quotidiano; neste sentido, corresponde a uma nova subida no nível de observação e explicação (na acepção de Reuchlin) no contínuo conceptualizável entre o funcionamento cognitivo molecular (processos neurofisiológicos, processos cognitivos) e a manifestação da inteligência no nível molar (comportamento em contexto).

Esta ascensão no nível de observação, inevitável a qualquer abordagem sistémica da inteligência, é, por isso mesmo, a que melhor caracteriza a avaliação da inteligência proposta no âmbito de uma outra teoria sistémica, a das Inteligências Múltiplas: Gardner defende que não tem sentido procurar aceder a medidas "puras" ou "directas" da inteligência – seja da "inteligência geral", da "inteligência musical", da "inteligência inter-pessoal", etc.[12] – porque essas "essências da alquimia cognitiva" (Gardner, 1999, p.207) não existem, antes derivam da tendência da ciência a criar terminologias, para de seguida lhes atribuir um sentido artificial de realidade (uma tendência cartesiana enraizada nas visões do mundo formista ou mecanicista, como se viu antes). O que importa avaliar, de acordo com Gardner, é a utilização que é feita das inteligências, separadamente ou em conjunto, para levar a cabo actividades e dar resposta a tarefas que são valorizadas pela sociedade em que se move um indivíduo, actividades e tarefas que presumivelmente as envolvem (o que desloca a avaliação para uma visão do mundo contextualista). Há, assim, que incidir na observação das pessoas em situações concretas de vida, aquelas em que, por exemplo, se revela a sua sensibilidade às necessidades ou aos motivos dos outros – observando a maneira como lidam com uma briga entre duas pessoas ou como convencem um superior hierárquico a mudar uma política, quando se pretenda avaliar a eficiência na esfera emocional e interpessoal. Um tipo de metodologia em pleno crescimento, e com enormes potencialidades para a avaliação da inteligência em contexto, é, de acordo com Gardner (1999, 2006), a da simulação: existe hoje tecnologia que permite a simulação de situações de carácter realista – "realidade virtual" – nas quais é também possível efectuar medidas das reacções e do desempenho muito exactas e até mesmo seleccionar os passos subsequentes da situação (modificar o contexto) em função das respostas ou reacções do examinado nas etapas anteriores – um formato de medida inspirado na "avaliação adaptativa", mas também passível de se adequar à chamada "avaliação

[12] Um argumento, recorde-se, também utilizado por Wechsler, a propósito da concepção de inteligência como manifestação da personalidade total, inseparável dos factores conativos, ou "não intelectivos".

dinâmica", se durante o desempenho for dado *feed-back* ou forem introduzidas ajudas e instruções, tendo em vista a avaliação do potencial de aprendizagem.

Este tipo de metodologias poderia ao mesmo tempo adaptar a avaliação em função das reacções emocionais e das manifestações motivacionais prévias do indivíduo avaliado, no sentido das supracitadas propostas de Raven (1989), e no quadro de uma visão do mundo selectivista, como ainda adequar a situação de observação em função da avaliação prévia do contexto habitual de vida do indivíduo, tendo em vista fazê-lo representar o melhor possível na situação de observação – algo que contribuiria para aumentar a "validade ecológica" das medidas, um conceito que tenderá a ganhar relevo na exacta medida em que a visão do mundo mecanicista, em que assenta grande parte das actuais técnicas de medida da inteligência, for substituída por uma visão do mundo contextualista. Mas este tipo de metodologias terá forçosamente de se apoiar em vastíssimos bancos de itens ou, mais ainda, na possibilidade de criação de itens no momento da própria resolução, duas ideias aparentemente arrojadas mas que fazem já parte de previsões actuais relativas à evolução da avaliação psicométrica para o novo século. Vale a pena, por isso, atender às principais linhas de evolução identificadas por grandes especialistas, desde as últimas duas décadas do século XX, no momento em que se aproximava o primeiro centenário da Psicometria e a transição de milénio: retêm-se três contributos, publicados com intervalos de cerca de uma década entre si – McReynolds (1982), Matarazzo (1992) e Embretson (2004).

Já no início da década de oitenta, McReynolds assinalava como duas tendências de evolução futura da medida psicológica a "avaliação sistemática do contexto" (que também designa de "avaliação ecológica") e a "avaliação interaccionista" (que toma em consideração simultaneamente os factores pessoais e situacionais) – em comum, a ênfase na avaliação contextualista – para o que preconizava a aplicação de técnicas como o "improviso sistemático" (p.128) ou o *roleplaying,* ao mesmo tempo que antecipava o crescimento da ênfase na avaliação de variáveis afectivas e motivacionais, em detrimento das variáveis cognitivas. Mas assinalava também, como terceira linha evolutiva de relevo, a profunda revisão da Teoria dos Testes, com o emergir de novas aplicações de desenvolvimentos metodológicos e tecnológicos: a aplicação da teoria da decisão à avaliação psicológica, por exemplo, para dar resposta ao problema da igualdade de oportunidades em selecção escolar e profissional; a divulgação da Teoria da Generalizabilidade (Cronbach, et al., 1972) para utilização sistemática nos estudos de precisão; e a implementação de novas técnicas de construção de testes. Embora menos explícito quanto a esta última inovação, não deixa de ser assinalável a sua pertinência numa época em que alguns dos postulados até então subjacentes à chamada Teoria Clássica dos Testes começavam a ser frontalmente questionados. Por exemplo, Raven em 1989, em publicação da *International Test Commission* atrás citada, discute um conjunto de princípios fundamentais – como, por exemplo, o de

que as distribuições de resultados dos testes devem seguir uma curva Normal ou gaussina, o de que a avaliação deve ser independente dos valores ou pode ser independente do contexto, o de que os testes devem ter elevada consistência interna, o de que a melhor maneira de obter descrições parcimoniosas das diferenças individuais é a partir de um pequeno número de variáveis, ou o de que a prática de proferir afirmações descritivas sobre os indivíduos não é científica, ao contrário da de efectuar discriminações inter-indivíduais a partir de um pequeno número de dimensões. Na mesma linha, nas últimas duas décadas, em consequência do enorme impacto da Teoria da Resposta ao Item, algumas "regras" clássicas subjacentes à medida modificaram-se radicalmente: Embretson e Reise (2000) listam dez regras clássicas que saem desafiadas por esta nova perspectiva de construção – por exemplo, a de que o erro-padrão da medida se aplica a todos os resultados numa determinada população (quando nos modelos de traço latente ele difere em função do nível de resultados, sendo maior nos resultados extremos e menor próximo da média); a de que quanto maior a extensão de um teste mais elevada a sua precisão (quando pela aplicação dos modelos de traço latente, ou TRI, testes mais curtos podem produzir medidas com maior precisão, desde que seja adequada a calibragem do nível de dificuldade dos itens ao nível de competência dos sujeitos avaliados); a de que apenas se pode obter estimativas das propriedades dos itens relativamente livres de erro se as observações forem efectuadas em amostras representativas (quando esse tipo de estimativas é possível nos modelos de traço latente a partir de amostras não representativas); a de que as propriedades das escalas de intervalos apenas se estabelecem se forem obtidas curvas Normais (quando tal se mostra possível pela aplicação de modelos de medida pertinentes). Uma das principais potencialidades da aplicação da Teoria da Resposta ao Item consiste precisamente na possibilidade de desenvolver bancos de itens apropriados à avaliação adaptativa, a forma de avaliação que, como se viu, se torna incontornável numa perspectiva contextualista, ou selectivista, da avaliação psicológica.

Na década de noventa, Matarazzo (1992) antecipa também a avaliação dos contextos de vida como uma das linhas de evolução futura, que levanta o problema de construir medidas "ecologicamente válidas" (p.1016), a par da avaliação das competências de adaptação e da medida da qualidade de vida. Ainda assim, a ênfase deste autor é colocada muito mais nas evoluções resultantes dos progressos das neurociências – a possibilidade de medição de índices da actividade cerebral durante a resolução de tarefas, como os testes de inteligência, e a imagiologia cerebral para estudo da estrutura do cérebro – razão por que cita numerosos estudos de correlatos neurofisiológicos dos resultados de testes de inteligência, muito designadamente dos resultados da *WAIS*. Uma bateria composta exclusivamente por medidas biológicas do funcionamento cortical que mostre predizer, tão bem quanto os actuais testes de inteligência, o sucesso escolar, o desempenho profissional ou mesmo outros aspectos da adaptação aos contextos comuns de vida, é vista por Matarazzo como um

instrumento viável e com potencial para o futuro da avaliação da inteligência, embora acredite que as técnicas tradicionais continuarão a ser utilizadas, paralelamente. Bem mais prudente relativamente a esta ênfase na avaliação neurobiológica da inteligência, pelo perigo evidente de reducionismo, mostra-se Sternberg que, embora não negue a possibilidade de recolha de evidências convergentes, de interesse tanto para a psicologia da inteligência como para as neurociências (Sternberg , 2005b) assinala o maior interesse de modelos "biopsicológicos" (Sternberg, Lautrey & Lubart, 2003b) que se situem na articulação entre conceptualizações *top-down* e *bottom-up* (pp.21-22), o que significa, por outras palavras, que evitem a omissão de qualquer classe de determinismos (ascendentes ou descendentes) ao contemplar múltiplos níveis de observação e explicação (na acepção de Reuchlin, 1999/2002).

Particularmente ricas e heurísticas são as "especulações" relativas ao "segundo século da medição da inteligência" efectuadas por Embretson em 2004, na medida em que articulam as tendências apontadas por outros autores com as espectaculares possibilidades tecnológicas que se abrem a partir das aplicações dos computadores e da *Internet* à avaliação psicológica. É curioso, e ao mesmo tempo fascinante, observar o quanto em apenas cerca de uma década evoluíram as perspectivas dos especialistas em avaliação psicológica quanto às possibilidades de aplicação da informática ao seu domínio (e imaginar o quanto poderão evoluir essas perspectivas nas próximas décadas): com efeito, em 1992, Matarazzo deixa como último apontamento, no seu artigo sobre o futuro da avaliação psicológica, uma referência ao desenvolvimento das aplicações dos computadores à interpretação de resultados dos testes, uma ideia que recorda ter aceite com relutância durante muito tempo, mas que acreditava na altura que viria a alcançar considerável sucesso num futuro mais ou menos próximo, na medida em que fosse possível efectuar cuidadosos estudos de validação de sistemas de avaliação altamente computorizados nos quais se articulassem, num todo com significado, informações de diversa ordem – sobre os contextos de vida objectivo e subjectivamente vivenciado (qualidade de vida), e sobre múltiplas funções cognitivas, conativas e afectivas medidas psicometricamente e neurofisiologicamente; apenas 12 anos mais tarde, são surpreendentemente mais amplas as aplicações das tecnologias informáticas antecipadas por Embretson (2004) para o domínio da avaliação psicológica, designadamente, quanto ao desenvolvimento de procedimentos de construção e aperfeiçoamento das técnicas de medida, quanto à natureza das tarefas da avaliação e quanto à própria delimitação do objecto da medida.

De acordo com esta autora, é possível prever, e para um futuro não muito distante, posto que muitos dos desenvolvimentos necessários estão já em curso, amplos sistemas informatizados e centralizados de avaliação psicológica acedidos via *Internet,* constituídos por gigantescas bases de itens, a que continuamente são acrescentados novos itens, a partir da calibragem relativamente aos

itens previamente existentes e do respectivo estudo metrológico. A revisão dos testes e a actualização das normas serão, então, parte integrante de um tal sistema e constituem processos em contínua actualização, à medida que novos protocolos e resultados vão sendo adicionados aos dados previamente disponíveis. Evidentemente, este tipo de sistema, que estará na base de formas de avaliação adaptativa, apenas se torna viável pela aplicação de procedimentos automáticos baseados em modelos de medida, no quadro da Teoria da Resposta ao Item – os quais conferem aos itens a propriedade da invariância[13] – e de princípios estatísticos de amostragem – que contemplam a ponderação sistemática de casos de modo a assegurar a representatividade das normas relativamente às populações de referência relevantes. No quadro destes sistemas centralizados de avaliação, os próprios estudos de validação procedem de forma automática, a partir da reunião contínua de respostas, reacções e tempos de execução, a par de outras informações demográficas – que viabilizam o estudo diferencial dos itens – e biográficas – relativas a critérios pertinentes à análise de correlatos dos resultados. Mais arrojada, mas já actualmente em pleno estudo, é a possibilidade de geração automática de novos itens: se um teste pode ser composto por itens automaticamente calibrados e também automaticamente seleccionados (em função do nível de competência do indivíduo avaliado ou de outros critérios pertinentes) para incorporarem uma avaliação adaptativa, não parece impossível a própria geração automática de itens a partir de programas de computador, sendo cada novo item uma variação de "modelos de itens" que previamente demonstraram potencialidades psicométricas adequadas. Nesta linha, pode-se até supor a possibilidade de geração instantânea de itens, ao longo da própria sessão de avaliação (novos itens criados automaticamente tendo por alvo determinadas propriedades metrológicas e a optimização da informação relativa ao sujeito examinado), o que leva a que se torne necessária não propriamente a calibragem prévia dos itens, mas a calibragem dos próprios modelos ou dos princípios subjacentes à geração dos itens. Estes programas de geração automática de itens fundamentar-se-ão, então, em modelos cognitivos relativos aos processos envolvidos na resolução dos itens, e na aplicação de modelos matemáticos para a predição dos tempos de execução e das propriedades metrológicas dos itens, em função das componentes cognitivas envolvidas. Uma importante vantagem destes sistemas automáticos de geração de itens consiste na possibilidade de minimizar a extensão da avaliação ao mesmo tempo que é maximizada a quantidade de informação recolhida sobre os modos e os níveis de funcionamento do examinado – por outras palavras, a possibilidade de reduzir a extensão dos testes, aumentando a sua precisão e maximizando a informação que fornecem.

[13] Para além da invariância dos parâmetros dos itens, que resulta de não serem enviesados pelas características das amostras (ver CAPÍTULO 4, pp.307-308), Embretson sublinha ser também imprescindível assegurar a invariância quanto aos itens específicos aplicados e quanto à sua posição relativamente às normas (Embretson, 2004, p.12).

Para lá da possibilidade de ligar directamente a medida à teoria psicológica (à teoria cognitiva), da ampliação da avaliação adaptativa, e da riqueza dos dados de validação em que se apoia a avaliação psicológica, Embretson antecipa ainda outras tendências, particularmente importantes se nos reportarmos às mais recentes propostas de Sternberg para a avaliação das inteligências prática e criativa: as tarefas de resposta aberta, como ensaios, resolução de problemas ou tarefas de completamento, cuja classificação requer o recurso a juízes humanos, irão de futuro poder ser classificadas também automaticamente, para o que estão já em ensaio programas informáticos, como o "*E-rater*" (Embretson, 2004, p.20) – que parte de uma amostragem de classificações efectuadas por juízes humanos a partir da qual determina, por métodos de regressão, os pesos a atribuir a diversos critérios, tendo em vista a optimização da predição, e de seguida classifica autonomamente os restantes ensaios a partir das ponderações identificadas. Por outro lado, a conceptualização do que pode constituir um item de um teste será muito ampliada, de novo a partir da aplicação de modelos necessariamente baseados na teoria psicológica, e abrangerá observações hoje dificilmente consideradas itens de testes – que poderão vir a incluir, por exemplo, tarefas de *role-playing* ou de improviso (McReynolds, 1982), ou mesmo simulações baseadas na realidade virtual, na linha das propostas de Gardner (1999) para a medida da inteligência. Acrescem ainda, a esta linha evolutiva, o muito mais importante papel que ganhará a avaliação de aspectos qualitativos, bem como a integração flexível, porque mediada por variáveis moderadoras, das avaliações quantitativa e qualitativa, e também a generalização da medida do desempenho perante modificação nas condições de observação – na linha do método hoje designado de "avaliação dinâmica" (Sternberg & Grigorenko, 2002).

Retomando as técnicas de medida da inteligência utilizadas na presente investigação, será então oportuno reportá-las a este quadro evolutivo, procurando desse modo entender não só o seu lugar na evolução da medida da inteligência, como algumas das suas potencialidades e limitações para o futuro da avaliação da inteligência. Relativamente à *WAIS-III*, este trabalho foi antes formalmente apresentado (Afonso, 2004a), tomando por referência uma análise anteriormente efectuada das tendências de evolução da medida da inteligência (Afonso, 2003a), às quais se acrescentam agora algumas das acabam de ser assinaladas. Assim, de seguida reporta-se a *WAIS-III* ao conjunto das mais importantes tendências de evolução identificadas nestes trabalhos:

1) Maior fundamentação teórica das medidas. Na base da maioria das linhas de evolução traçadas, de maneira mais ou menos explícita, assinala-se a muito maior ligação entre a medida e a teoria psicológica – os modelos de medida, os modelos de itens, os modelos cognitivos, etc. remetem para um quadro de referência teórico que funciona como ponto de partida para a construção das medidas (numa lógica *top-down*), em contraste com muitas

técnicas de avaliação do passado, designadamente factoriais, construídas a partir das evidências empíricas das diferenças individuais (numa lógica *bottom-up*). A *WAIS*, a par das restantes escalas de Wechsler, constitui uma excepção no que aos testes "clássicos" diz respeito, por se apoiar explicitamente, e desde a sua primeira edição, numa conceptualização da inteligência humana – trata-se, deste ponto de vista, de um instrumento construído numa lógica *top-down*. Uma leitura dos manuais da mais recente edição da *WAIS* (Wechsler, 1997a,b) revela que continuam a situar o fundamento teórico da prova por referência às ideias originais de Wechsler, mantendo assim a sua fidelidade relativamente a uma conceptualização tradicional da inteligência, aliás, como antes referido, ainda hoje aceite por muitos autores e apoiada por muito importante corpo de investigação. Acrescentam-se, apesar disso, e pela primeira vez, outros conceitos em articulação com o de inteligência, designadamente os de aprendizagem e de memória, de onde decorre uma maior explicitação da natureza do construto. Algumas ambiguidades persistem, contudo, no enquadramento conceptual: a ênfase no carácter multifactorial da estrutura das medidas, tanto na análise factorial exploratória como na confirmatória, dando relevo às evidências que apoiam os índices factoriais mais do que às que apoiam os índices globais, em particular o QI de Escala Completa, não sendo retiradas, no entanto, as devidas implicações para a natureza do construto e para a interpretação dos resultados globais (QIs); a possibilidade de opção pelos Índices Factoriais, ao invés dos QIs, de certa forma desvirtuando a avaliação da inteligência tal como Wechsler a concebeu; a articulação num mesmo nível de variáveis de natureza distinta, compreensão verbal e organização perceptiva, variáveis "traço", a par com memória de trabalho e velocidade de processamento, variáveis de "mediação cognitiva" ou de "processo", tratadas então como variáveis traço; a ausência de articulação explícita entre a estrutura das medidas da prova e os mais aceites modelos de organização das aptidões humanas, como, por exemplo, a Teoria dos Três Estratos de Carroll (2003) ou, de modo mais genérico, o modelo C-H-C. Assim, embora haja uma preocupação maior com a fundamentação conceptual das medidas, explicitamente assinalada no Manual Técnico (Wechsler, 1997b, p. 12), ela não é nem suficientemente clara, nem completamente rigorosa.

2) Ampliação do conceito de inteligência. Uma segunda tendência que emerge na análise da evolução da medida da inteligência consiste na ampliação do próprio conceito, que vem a abranger aspectos conativos e afectivos e tende a alargar os seus contornos e a assumir uma natureza multifacetada, por comparação com a noção tradicional de inteligência geral. Ao remeter o enquadramento teórico para as propostas originais de Wechsler, não se

observa na *WAIS-III* uma ampliação do conceito de inteligência semelhante à que caracteriza algumas propostas actuais. É, contudo, notável, como assinalado no CAPÍTULO 3, a forma ampla como a inteligência é definida desde o início pelo autor, ao reconhecer, por exemplo, o papel dos "factores não intelectivos" ou "conativos" no desempenho cognitivo (Wechsler, 1975), factores hoje fulcrais nas novas teorias da inteligência (Salovey & Mayer, 1990; Sternberg, 1996; Gardner, 1999) e ao entendê-la como manifestação da personalidade total. Mesmo assim, de acordo com Sternberg (2002a,b; 2003g), entre outros, nos testes tradicionais de inteligência, como a *WAIS-III*, um factor geral emerge porque a gama de aptidões avaliada é muito estreita e incompleta.

3) Construção de novos testes dirigidos a novos construtos, provenientes da psicologia cognitiva. Como se viu, os modelos teóricos que se prevê virem a estabelecer o quadro de referência para o desenvolvimento das medidas de inteligência do futuro são os que provêm da psicologia cognitiva, posto que são os que postulam os processos mentais subjacentes ao desempenho nos testes de inteligência, permitindo que a medida psicológica escape ao pragmatismo clássico que toma por suficiente a avaliação quantitativa dos produtos da actividade mental. Ao avaliar em que medida esta tendência está presente na *WAIS-III,* é certo que se verifica que novos subtestes foram introduzidos, embora talvez não com o propósito claro de ampliar a gama de funcionamento avaliada, ou de medir novas variáveis. É antes sublinhada nos manuais a intenção do reforço da medição de variáveis tradicionalmente identificadas na análise factorial – como a memória de trabalho e a velocidade de processamento – sendo apenas ligeiramente acentuado o peso das variáveis de processo cognitivo no conjunto da avaliação, mesmo estas avaliadas acima de tudo quanto ao produto do seu funcionamento (como dimensões).
4) Generalização da avaliação interactiva, adaptativa e contextual. Esta é talvez uma das tendências de evolução mais marcantes e mais consensuais entre os autores que se debruçam sobre o futuro da medição psicológica: a avaliação do contexto ("avaliação ecológica"), a avaliação em contexto ("avaliação interaccionista") e a avaliação perante modificação das condições do contexto ("avaliação dinâmica") representam diversas facetas de uma mesma ênfase na contextualização da medição psicológica ou, por outras palavras, assinalam uma mudança mais fundamental de uma visão do mundo mecanicista para uma visão do mundo contextualista (Pepper, 1942), em avaliação psicológica. Os procedimentos de aplicação adaptativos, de futuro mais viabilizados ainda pelos progressos tecnológicos em curso, são imprescindíveis à avaliação na perspectiva interaccionista, ecológica e dinâmica. A *WAIS-III* não é certamente uma técnica de

avaliação adaptativa, embora tenha havido nesta edição algum cuidado com a o estabelecimento mais rigoroso, e também mais flexível, dos procedimentos de inversão (em caso de insucesso nos primeiros itens aplicados) ou de paragem (após um determinado número de insucessos consecutivos). Mas a larga maioria dos itens aplicados é semelhante para todos os sujeitos, de onde decorreu, neste estudo, que alguns se tenham mostrado mesmo desajustados ao nível de competência das amostras examinadas. Por outro lado, uma das tendências de evolução que não parece ter orientado a revisão mais recente da *WAIS* é a que se refere à adopção de procedimentos dinâmicos de aplicação. Pelo contrário, há até evidência de um esforço no sentido de estabelecer ainda com mais detalhe todas as regras de aplicação estandardizada e de quantificação das respostas, o que acentua o carácter de "avaliação estática". Por fim, na *WAIS-III*, por ser uma prova de aplicação individual, existe alguma possibilidade de reportar as respostas do examinado ao contexto imediato da avaliação ou, eventualmente, a partir de uma adequada anamnese, a algumas informações relativas ao seu contexto actual de vida. Este não é, contudo, objecto de avaliação sistemática, nem é modificado em função da acção do sujeito durante a resolução do teste, como acontece numa avaliação interactiva; pelo contrário, devido ao reforço da avaliação "estática" em detrimento de "dinâmica", a terceira edição da *WAIS* não parece ter apostado em integrar esta tendência de evolução, persistindo o seu carácter de técnica de avaliação centrada nos aspectos do mundo interno do indivíduo (ver *Loci da Inteligência Humana,* CAPÍTULO 1, p.32) e radicada numa visão do mundo mecanicista.

5) Maior ênfase no diagnóstico cognitivo do que na avaliação do desempenho. A ideia de que a avaliação psicológica deve incidir na "compreensão" do funcionamento mais do que na "explicação" do funcionamento (baseada nos traços latentes) (uma mudança que não abandona, ainda assim, uma heurística elementarista (Overton, 2002), terá conduzido ao crescimento da ênfase na avaliação de aspectos qualitativos do desempenho, para lá dos quantitativos, e na leitura da configuração dos dados de avaliação mais do que na pesquisa dos factores explicativos desses dados. Algumas inovações da 3ª edição da *WAIS* são subtis mas mesmo assim reveladoras de alguma mudança no sentido de uma avaliação centrada na compreensão do funcionamento ou "diagnóstico cognitivo" (Huteau & Lautrey, 1999), mais do que na avaliação quantitativa do funcionamento por referência à população. Esta preocupação manifesta-se na tentativa de identificação de áreas de rendimento ou de potencial, e de áreas de défice ou modos de processamento menos eficazes, por exemplo, pela introdução de procedimentos opcionais (como no subteste de

Código:Dígito-Símbolo), ou pela maior aposta na avaliação de variáveis com carácter funcional, como a memória de trabalho e a velocidade de processamento. A análise de perfis e a aposta na comparação sistemática de resultados e índices, viabilizadas por um completo conjunto de tabelas para interpretação de discrepâncias de resultados, revelam o acentuar desta perspectiva intra-individual de análise.

6) Maior ênfase na abordagem idiográfica do que na abordagem nomotética. Esta maior atenção às diferenças intra-individuais do que inter-individuais implica o declínio da abordagem nomotética da avaliação e o afirmar da avaliação idiográfica. A acentuação do carácter idiográfico da medida da inteligência na *WAIS-III* está de algum modo presente nas modificações referidas no parágrafo anterior, ao que acresce a mudança no procedimento de derivação dos resultados dos subtestes, que toma por comparação os resultados do respectivo grupo etário, diferentemente do procedimento adoptado nas anteriores edições da prova, em que era utilizado um grupo de referência de máximo rendimento médio na população adulta (20-34 anos); esta alteração acentua a leitura desenvolvimentista e a comparação intra-individual dos resultados ao nível dos subtestes. Ainda assim, todas as análises, mesmo as intra-individuais, começam por comparar sempre o indivíduo com uma norma, o que reporta ainda a análise intra-individual à comparação inter-individual. Por outro lado, as informações de carácter qualitativo, embora sejam desde sempre incluídas na interpretação dos resultados da *WAIS*, não são recolhidas de acordo com grelhas de observação formais, pelo que não são sistemáticas nem incidem, eventualmente, nos aspectos mais pertinentes do ponto de vista teórico. Um aspecto particular da mudança de ênfase da avaliação nomotética para a idiográfica consiste na tendência a substituir o formato de resposta múltipla, comum a muitas técnicas de avaliação cognitiva, pelo de resposta aberta. Há que enaltecer a inclusão de tarefas de resposta aberta que caracteriza as escalas de inteligência de Wechsler desde as suas primeiras edições, algo que se justifica pelo contexto clínico da sua origem e se viabiliza pelo carácter individual da aplicação; a utilização de critérios de classificação e o estudo da consistência entre as pontuações atribuídas por diferentes avaliadores são metodologias já tradicionais nestas provas, cujas potencialidades, evidenciadas pelos respectivos estudos metrológicos, só podem servir de estímulo à maior generalização deste formato de resposta na avaliação da inteligência humana.

7) Modificação nos propósitos da avaliação psicológica. Uma tendência de evolução decorrente das duas mudanças de ênfase que acabam de ser tratadas consiste na maior atenção à promoção do bem-estar, apostando nas potencialidades e promovendo a

correcção dos défices, do que no mero diagnóstico da deficiência. Esta mudança é evidentemente enquadrada por uma modificação mais fundamental, de uma "psicologia da patologia" (centrada no despiste da deficiência) para uma "psicologia positiva" (centrada na promoção da adaptabilidade), na qual ganham relevo conceitos como o de bem-estar psicológico ou de qualidade-de-vida (Lopez & Snyder, 2004). Pela maior ênfase na avaliação idiográfica e a crescente atenção à compreensão do funcionamento através do "diagnóstico cognitivo", a utilização da *WAIS-III* mostra-se útil no âmbito de uma intervenção psicológica "positiva".

8) Implantação de novas técnicas estatísticas de construção e estudo metrológico. Uma das mais importantes evoluções no domínio da avaliação psicológica diz respeito, como se viu, ao aperfeiçoamento dos métodos estatísticos subjacentes à construção e estudo metrológico das técnicas de avaliação, uma área em que se esperam, aliás, importantes desenvolvimentos para as próximas décadas. É inquestionável o esforço de actualização das metodologias de estudo metrológico e de aperfeiçoamento técnico da *WAIS-III*, e a apresentação extensiva de resultados empíricos muito sólidos que preenche boa parte do Manual Técnico. Acrescente-se, a propósito, que este é mesmo modelar, pela sua excelência em termos de riqueza de conteúdo, de clareza de organização e de rigor técnico. Na generalidade, contudo, a Teoria Clássica dos Testes tende a preponderar e as técnicas de análise só pontualmente fazem apelo a conceitos ou índices enquadrados pela Teoria da Resposta ao Item, por exemplo, o que mais não faz do que reforçar a natureza não adaptativa e estática da avaliação proporcionada pela *WAIS-III*. Ao que parece, embora sejam reconhecidas algumas potencialidades das mais recentes perspectivas de construção e estudo metrológico, estas não são adoptadas, muito possivelmente porque essa opção implicaria uma modificação substancial da estrutura e talvez do conteúdo da *WAIS*.

9) Acentuação da interdisciplinaridade. A interdisciplinaridade, uma das tendências actuais de evolução da investigação no domínio da inteligência (Miranda, 2001; Sternberg, Lautrey & Lubart, 2003b),emergiu também no levantamento das tendências de evolução da medida psicológica para as próximas décadas: a aplicação dos progressos das neurociências, da imagiologia cerebral, da genética, bem como das tecnologias de informação e da matemática à avaliação psicológica, em particular à medida da inteligência, constituem alguns exemplos desta tendência. Numerosos estudos têm sido efectuados procurando determinar os correlatos neurofisiológicos do desempenho na *WAIS,* algo que parece ampliar as possibilidades de avaliação da inteligência para lá dos instrumentos

convencionais. Esta tendência marca de algum modo a nova edição da *WAIS*, pelas claras ligações que são estabelecidas com domínios de investigação e de aplicação que ultrapassam o da clássica avaliação diferencial: por exemplo, a aplicação da prova em neuropsicologia é empiricamente fundamentada em diversos grupos clínicos com lesão cerebral localizada e com doença neurológica diagnosticada. Mas ficam ainda em aberto, talvez por prudência, muitas outras possibilidades de articulação com outros domínios, numa óptica de interdisciplinaridade que seria certamente cara a Wechsler.

A *WAIS-III* constitui, em suma, uma técnica que bem pode ser classificada de diferencial, ao manter um conjunto de características dos métodos tradicionais de avaliação da inteligência, mas que contém, ao mesmo tempo, potencialidades para poder evoluir no sentido de alguns vectores actuais de desenvolvimento da medida da inteligência, ou mais genericamente, da avaliação psicológica. Este sentido de actualidade das escalas de Wechsler é provavelmente justificado pelo quadro de referência teórico e metodológico que o seu autor esboçou, desde há mais de seis décadas, o qual, como se viu (CAPÍTULO 3) foi marcado pela sua própria evolução de uma visão do mundo mecanicista para uma visão do mundo contextualista e, mais tarde, selectivista (na acepção de Pepper, 1966).

Chegou o momento de proceder a uma análise equivalente à anterior, desta feita incidindo sobre o *STAT-R (H)* (2004). Retomam-se, para o efeito, os mesmas nove tópicos de síntese da evolução da medida da inteligência:

1) Maior fundamentação teórica das medidas. O *STAT-R (H)* representa, como se viu, um dos melhores exemplos de um método de avaliação da inteligência criado a partir de uma teoria da inteligência – a Teoria Triárquica da Inteligência Humana – e como operacionalização de um novo conceito – o de Inteligência Funcional. A este respeito marca, por isso mesmo, uma mudança importante no domínio da medida da inteligência, sobretudo se o compararmos com a maioria dos testes colectivos, ou de papel-e-lápis, métodos em geral construídos numa lógica *bottom-up*. Contudo, algumas incoerências ou ambiguidades chamam a atenção: primeiro, o facto de tratar-se de uma prova de formato tradicional, e de avaliação estritamente quantitativa, o que não se mostra coerente com a concepção que pretende representar, em que a inteligência é entendida como equilíbrio entre diversas formas de processamento mental qualitativamente diferenciadas; segundo, o facto de ser uma prova de aplicação estandardizada, que reduz a possibilidade de avaliação contextualizada, sublinhada pela subteoria contextual; terceiro, o facto de se tratar de uma técnica de resposta por escolha múltipla, reduzindo a possibilidade de manifestação espontânea, supostamente útil para a avaliação da inteligência criativa, no quadro da subteoria experiencial; quarto, o apuramento de um resultado global, que

contraria a noção da inteligência como equilíbrio e a convicção na independência relativa das três formas de inteligência que o *STAT-R (H)* supostamente avalia. Enfim, sem pretender esgotar a lista das incoerências entre a operacionalização e a teoria, posto que muitas delas serão abordadas nos tópicos seguintes, parece desde já evidente que a estrutura e natureza do teste não se afigura completamente consistente com o modelo teórico que procura operacionalizar.

2) Ampliação do conceito de inteligência. Esta constitui uma das tendências sem dúvida bem representada no *STAT-R (H)*: trata-se de uma técnica de avaliação que procura de maneira muito explícita alargar o espectro do funcionamento mental abrangido pela avaliação da inteligência, ao incluir a avaliação das inteligências prática e criativa para além da inteligência analítica, medida pelos testes tradicionais de inteligência. O estudo metrológico, quer internacional quer das amostras do presente estudo, não oferece, infelizmente, resultados completamente claros ou inequívocos relativamente a este alargamento: por um lado, o facto de se manifestar a presença de um factor comum a todas as formas de inteligência acaba por sugerir a interpretação das medidas das inteligências prática e criativa como novas formas de medir a mesma inteligência geral, desde há muito medida pelos testes de natureza analítica; acresce que a elevada correlação entre a parte Criativa e as medidas da inteligência fluida sugere alguma sobreposição destes dois construtos; por outro lado, a baixa precisão das medidas contribui para levantar dúvidas sobre a adequação e a aplicabilidade prática desta forma de operacionalização das inteligências prática e criativa, o que terá motivado Sternberg a ensaiar racionais alternativos para a sua medição.

3) Construção de novos testes dirigidos a novos construtos, provenientes da psicologia cognitiva. De novo, encontra-se no *STAT-R (H)* um excelente exemplo de aplicação de mais esta tendência de evolução da medida da inteligência: a subteoria componencial, resultado mais directo da investigação cognitiva das aptidões que ocupou Sternberg nos primórdios dos seus trabalhos, está presente, de acordo com o autor, em toda a bateria: no domínio analítico, as metacomponentes, as componentes de execução e as componentes de aquisição de conhecimento aplicam-se a tarefas abstractas e familiares; no domínio prático, aplicam-se a tarefas relativamente familiares mas de conteúdo concreto, ligado a problemas não académicos; e no domínio criativo, aplicam-se a tarefas que exigem adaptação à novidade, ou seja, capacidade de lidar eficazmente com tarefas não familiares. Ainda assim, embora estas componentes, como se viu antes (CAPÍTULO 3) sejam consideradas pelo autor como os mais fundamentais e universais processos subjacentes à

inteligência, o *STAT-R (H)* não possibilita, e nem sequer pretende, discernir o nível ou o modo de funcionamento destas componentes. Assim, apesar de baseado na investigação da cognição humana, o *STAT-R (H)* mais pretende avaliar o sucesso na resposta a problemas de natureza diversa, que envolvem sempre uma combinação não conhecida nem averiguada de componentes de processamento de informação, no que é uma técnica bem mais próxima do que se propunha inicialmente da avaliação convencional da inteligência (pouco preocupada com o funcionamento ou com a combinação de processos mentais envolvidos na resolução dos itens, antes centrada na avaliação da eficácia da aplicação desses processos à resolução de determinadas classes de problemas). Por outras palavras, ao assemelhar-se às técnicas tradicionais de avaliação da inteligência, o *STAT-R (H)* não corresponde, como pretende o seu autor, a esta tendência de evolução, e radica numa visão do mundo mecanicista – porque avalia, afinal, a inteligência do ponto de vista do mundo interno do indivíduo (ver *Loci da Inteligência Humana,* CAPÍTULO 1, p.32).

4) Generalização da avaliação interactiva, adaptativa e contextual. Talvez paradoxalmente, posto que a Teoria Triárquica inclui uma subteoria contextual, e afirma-se como representando uma postura contextualista de investigação da inteligência, o *STAT-R (H)* constitui um meio bem convencional de avaliação da inteligência, no respeitante a este critério, ao assumir o tradicional formato de teste de papel-e-lápis e de aplicação estandardizada. Não está, por conseguinte, em causa, nem a avaliação interaccionista, posto que não existe interacção complexa com o meio ao longo da resolução, nem a avaliação adaptativa, posto que o teste é igual para todos os indivíduos avaliados, nem a avaliação dinâmica, posto que não existe nenhuma oportunidade de treino ou de aprendizagem no decurso da realização da prova, e nem sequer a avaliação contextual, posto que, como prova de aplicação colectiva, permite pouco perceber o efeito dos factores contextuais imediatos no desempenho, não tomando também em consideração os factores contextuais da vida presente do examinado na interpretação dos resultado. Não se pode deixar de assinalar o quanto é surpreendente alguma desarticulação entre o *STAT-R (H)* e as perspectivas teóricas e metateóricas que o seu autor afirma estarem na base da sua construção. Mas é importante sublinhar a tentativa de maior contextualização da avaliação presente quer nos itens-ensaio quer, de maneira mais expressiva, nas técnicas propostas mais recentemente para a avaliação da inteligência prática e da inteligência criativa.
5) Maior ênfase no diagnóstico cognitivo do que na avaliação do desempenho. Esta é uma ênfase repetidamente assinalada pelo autor do *STAT-R (H)*: não se pretende avaliar a inteligência como um todo, um resultado desprovido de aplicação na promoção do bem-

estar ou da adaptação do indivíduo; pelo contrário, considera-se mais adequada, deste ponto de vista, a adequada identificação das potencialidades e dos défices, que permita capitalizar as primeiras e compensar ou corrigir os segundos. A demonstração da existência de correlações expressivas entre os três domínios e o emergir de uma estrutura factorial, nas amostras do presente estudo, dando mais relevo à distinção entre áreas de conteúdo, ou entre noções clássicas de inteligência (fluida e cristalizada, por exemplo), do que entre domínios de processamento coloca algumas reservas, contudo, à utilidade das distinções entre os resultados do *STAT-R (H)* apoiados na Teoria Triárquica.

6) Maior ênfase na abordagem idiográfica do que na abordagem nomotética. Apesar desta aparente atenção às diferenças intra-individuais, o *STAT-R (H)* não se mostra inteiramente coerente com ela, seja pela ênfase nos resultados quantitativos e omissão quanto à obtenção de informações qualitativas sobre o desempenho, seja pela limitada informação intra-individual que efectivamente proporciona, seja acima de tudo pela fragilidade, atrás comentada, das distinções relativas ao funcionamento cognitivo que introduz. Mas é importante acrescentar que a adição de itens-ensaio, itens de resposta aberta que permitem uma apreensão mais qualitativa do processamento mental e, mais tarde, de novas tarefas criadas para a avaliação das inteligências prática e criativa, mais viradas para a avaliação do desempenho ou para a avaliação em contexto, abrem maiores possibilidades à avaliação de tónica idiográfica.
7) Modificação nos propósitos da avaliação psicológica. Esta modificação quanto aos propósitos de avaliação da inteligência é parte integrante da Teoria da Inteligência Funcional; no entanto, pelo que acabou de ser comentado relativamente ao *STAT-R (H)*, o formato estandardizado e de resposta múltipla, bem como a ênfase na avaliação quantitativa, mais do que na descrição qualitativa do desempenho, não se mostra de todo favorável à obtenção de informações úteis à promoção do bem-estar e da adaptação, no respeito pela idiossincrasia individual, mais acentuam a comparação inter-individual e a adopção de uma perspectiva convencional e nomotética de medida da inteligência..
8) Implantação de novas técnicas estatísticas de construção e estudo metrológico. Os estudos de validação da Teoria Triárquica que utilizam, entre outros instrumentos de avaliação, o *STAT-R (H)* têm vindo a evidenciar crescente recurso a técnicas sofisticadas de análise de dados. Exemplo claro encontra-se nas mais recentes publicações sobre o projecto *Rainbow* (ver, por exemplo, Sternberg & Rainbow Project Collab., 2006). Mas importa assinalar que as referências específicas ao estudo metrológico do *STAT-R (H)* ocupam em geral uma parte diminuta da apresentação dos resultados e comportam

algumas imprecisões importantes: por exemplo, como evidência da consistência interna é frequente, como é o caso no estudo atrás citado (p.332), ser apresentado o coeficiente relativo ao conjunto dos 45 itens, o que é absolutamente incoerente com a noção de que o *STAT-R (H)* mede três formas *distintas* de inteligência – a única justificação para um tal procedimento reside, talvez, em que esse coeficiente atinge em geral um valor relativamente elevado, superior a .70, o que legitima, apenas aparentemente, a utilização do *STAT-R (H)* em investigação. Por outro lado, os coeficientes menos que satisfatórios dos resultados dos domínios são justificados pelo "baixo número de itens" (p.332), uma explicação que se mostra pertinente para os resultados de cada parte (apenas 5 itens) mas que é já discutível quanto à extensão de cada domínio (15 itens)[14].

9) Acentuação da interdisciplinaridade. Embora a Teoria Triárquica, pelo seu carácter plurimetafórico, implique, tal como qualquer teoria sistémica, a integração dos contributos de uma multiplicidade de domínios de conhecimento, o *STAT-R (H)* não reflecte em si esta interdisciplinaridade, mais radicando na concepção diferencial da avaliação da inteligência. A perspectiva de Sternberg relativamente à investigação da inteligência sublinha a absoluta necessidade de convergência de diferentes contributos: "as abordagens biológicas para a compreensão da inteligência são de extrema utilidade, à semelhança de outras abordagens, como a psicométrica, a cognitiva, a cultural, a educacional, para nomear apenas algumas. A melhor maneira de compreender a inteligência é através de operações convergentes […] utilizando uma variedade de técnicas empíricas que, espera-se, convirjam num conjunto coerente de resultados que demonstrem transcender os métodos usados para os obter" (Sternberg, 2005b, pp.243-244). Deste ponto de vista, o *STAT-R (H)* corresponderá, pelo que se viu, a uma técnica empírica representativa da abordagem psicométrica, ainda que inspirada também na abordagem cognitiva da inteligência.

Quando reportadas aos principais vectores de desenvolvimento actual da avaliação da inteligência, qualquer das técnicas diferenciais utilizadas no presente estudo apresenta potencialidades e limitações. Quanto à *WAIS-III*, as potencialidades decorrem em grande parte do conceito amplo e integrativo de inteligência em que se apoia o qual, embora não tenha sido actualizado relativamente às propostas originais do seu autor, desde a origem se enquadrou numa visão do mundo contextualista, ou mesmo selectivista, antecipando a ênfase que só mais tarde viria a afirmar-se nas concepções sistémicas de inteligência. O facto de ser uma bateria de aplicação individual abre ao mesmo tempo a possibilidade de adequação, desde que devidamente fundamentada, à avaliação interactiva,

[14] Alguns subtestes da *WAIS-III*, mesmo com número equivalente ou menor de itens, permitem obter coeficientes superiores aos dos domínios do *STAT-R (H)*.

designadamente à avaliação dinâmica, uma metodologia ensaiada com êxito em outras escalas de Wechsler, como a *WPPSI-R* (Seabra-Santos, 2000), ou à avaliação adaptativa, desde que a estrutura e a forma de aplicação sejam modificadas em conformidade. O facto de fornecer resultados situados em níveis distintos de observação e explicação, aliado à possibilidade de proporcionar também dados qualitativos relativos ao desempenho numa gama diversa de situações e tarefas, e à possibilidade de averiguação do desempenho através de tarefas opcionais, concede à *WAIS-III* o carácter de uma técnica estrutural de avaliação da inteligência (não simplesmente holística, nem, de todo, elementarista). O reforço do carácter estático da aplicação, consequência do investimento na estandardização das condições de observação, constitui uma linha de evolução contrária à que seria desejável, face às futuras tendências da avaliação psicológica – maior relevo da avaliação do contexto e em contexto – e ao assumir, por inerência, da visão do mundo contextualista. A visão do mundo selectivista, uma elaboração da visão contextualista mas que lhe acrescenta a intencionalidade característica dos sistemas de elevada complexidade, apelaria não só a maior ênfase na avaliação funcional do indivíduo em contexto mas ainda à exploração mais formal dos factores "não-intelectivos" ou "conativos", que já Wechsler reconhecia influenciarem o desempenho cognitivo, mas os quais são observados de maneira pouco sistemática e objectiva na *WAIS*. Uma das implicações para o futuro desenvolvimento desta técnica diferencial seria incluir uma grelha de observação ou uma medida (ou conjunto de medidas) incidindo nos aspectos motivacionais e emocionais manifestos durante a avaliação.

O *STAT-R (H)* ao representar uma tentativa de alargamento da avaliação da inteligência e consistir na operacionalização de uma teoria previamente edificada, constitui uma técnica bem exemplificativa das actuais perspectivas de construção de métodos de avaliação, muito designadamente, no domínio da inteligência. Por outro lado, o facto de as suas partes constituintes se apoiarem em racionais que se fundamentam em conceitos e modelos oriundos da psicologia cognitiva consubstancia, a nível da medida, a integração entre os paradigmas de investigação diferencial e informacional que moveram o autor nos primórdios dos seus trabalhos. A maior fragilidade do *STAT-R (H)*, aliás, da qual se admite resultarem em alguma parte as limitações detectadas no estudo metrológico (nomeadamente as que dizem respeito à estrutura interna das medidas), decorre talvez de ser uma técnica radicada numa visão do mundo mecanicista (ao articular os referidos paradigmas) que, no entanto, pretende constituir operacionalização de uma teoria radicada numa visão do mundo contextualista/selectivista. De que modo no *STAT-R (H)* aparece operacionalizada a noção de que a inteligência só pode ser avaliada em função do contexto? E de que modo no *STAT-R (H)* aparece operacionalizada a noção de que a inteligência deve ser avaliada em função dos valores e objectivos pessoais de cada indivíduo? Ao contrário do que se poderia esperar, atendendo à tão veemente crítica

de Sternberg em relação à tradicional medição da inteligência, o *STAT-R (H)* é uma técnica ainda centrada na avaliação de aspectos do mundo interno do indivíduo, mesmo quando está em causa a medida da inteligência prática e da inteligência criativa: no primeiro caso, pela exclusiva representação do processo de "adaptação" (modificação do próprio para responder às pressões do meio) e concomitante omissão da avaliação da acção sobre o meio ou, eventualmente, da delimitação de critérios para a selecção de um novo meio; no segundo caso, pela utilização de tarefas fechadas, de resposta por escolha múltipla, na avaliação da "capacidade" para lidar com a novidade – uma noção próxima afinal do já bem conhecido conceito de inteligência fluida. Acresce que o carácter de avaliação estática, decorrente da natureza estandardizada e não adaptativa da aplicação, mais acentua a aproximação à concepção tradicional de medida em que a inteligência é entendida e avaliada enquanto atributo do indivíduo – uma constatação sem dúvida surpreendente se recordarmos a convicção do autor em que "a inteligência não está simplesmente na cabeça" [*Intelligence is not just inside the head*, (Sternberg, 2002b)]. Sem dúvida, a evolução recente no que toca a construção de novas técnicas de avaliação por parte de Sternberg e seus colaboradores procura suprir em parte este desajuste entre teoria e operacionalização, através da substituição de itens de resposta múltipla por itens-ensaio, ou por itens de realização ou de desempenho, e através da introdução da possibilidade de opção, por parte do examinado, relativamente aos itens para os quais sente maior motivação. Estas formas de avaliação mantêm, contudo, algumas limitações clássicas de técnicas avaliação da inteligência, como seja o facto de envolverem em larga escala a utilização da linguagem, o que provavelmente contamina a avaliação dos domínios prático e criativo de factor verbal e condiciona as possibilidades de generalização a outras formas de manifestação desses aspectos da inteligência.

A avaliação da inteligência no quadro do paradigma sistémico terá necessariamente de romper de maneira mais radical com as formas tradicionais de avaliação, posto que corresponde mais do que a uma mudança de paradigma a uma profunda mudança do próprio objecto de avaliação. Não se trata de defender que as técnicas tradicionais de medida estão ultrapassadas ou deviam ser simplesmente substituídas por novas técnicas; bem antes pelo contrário, como se viu, numa técnica tradicional como a *WAIS-III* foram detectadas enormes potencialidades para a avaliação da inteligência numa óptica contextualista e sistémica. Trata-se antes de argumentar que enquanto se persistir num formato de avaliação tradicional mais não se fará do que construir técnicas que, tal como os vários subtestes das baterias de Wechsler, medem de diferentes maneiras sempre a mesma inteligência (ou as mesmas formas de inteligência) – a que se manifesta nas diferenças individuais em tarefas estandardizadas e emerge, por isso, nas correlações entre resultados quantitativos apurados nas respostas a essas tarefas. As várias técnicas de avaliação limitam-se a repartir de maneiras diversas a mesma variância,

relativa a um mesmo nível de observação e explicação (o nível molar dos factores do mundo interno do indivíduo).

O paradigma sistémico de investigação e de avaliação da inteligência terá de assumir, de maneira explícita e inequívoca, o enquadramento contextualista que lhe é inerente, e incidir na interacção entre as características do sujeito (do mundo interno) e as características do contexto (do mundo externo). Note-se que a definição do aspecto contextual da inteligência, tal como Sternberg a apresenta nalgumas exposições da Teoria Triárquica (por exemplo, Sternberg, 1985a) como a relação da inteligência com o mundo externo do indivíduo, ou como envolvendo a modificação do indivíduo para se ajustar ao contexto (adaptação), a modificação do contexto para o adaptar ao indivíduo (modelação) ou a selecção de um novo meio por parte do indivíduo (selecção), mantém o centro de gravidade da inteligência no indivíduo, deixando ao meio o papel de cenário da acção – o que terá conduzido a operacionalizações que reproduzem, em boa parte, as medidas tradicionais da inteligência centradas na inteligência como atributo do indivíduo e mais representativas das visões do mundo formista e mecanicista do que da contextualista. Mas, como sublinha Van Geert (2003, p.199), "uma vez que a acção adaptativa é por natureza dependente do contexto, a solução mais económica para determinar a origem da acção adaptativa é por referência a um mecanismo ou mecanismos que consistam numa combinação dos dois factores envolvidos, designadamente, o organismo e os seus contextos de acção". Não se trata de entender a inteligência apenas a partir do indivíduo ou da sua acção sobre o meio (objecto privilegiado dos paradigmas biológico, diferencial, construtivista e informacional), nem apenas a partir do meio ou da sua acção sobre o indivíduo (objecto privilegiado dos paradigmas antropológico e sociológico), mas sim através do *processo*, dinâmico e temporal, de acção recíproca (que deverá ser, este sim, o objecto privilegiado do paradigma sistémico). Neste sentido, é igualmente pertinente falar de "indivíduos inteligentes" como de "contextos inteligentes" – que outro objectivo para a organização social, económica e política senão o de tornar os contextos mais favoráveis (leia-se "mais inteligentes", "mais adaptados") à implementação dos objectivos e dos valores dos indivíduos e das comunidades humanas, em face dos circunstancialismos presentes em determinado período da História da humanidade e do desenvolvimento da espécie?[15] Esta perspectiva, com toda a evidência inscrita numa metateoria relacional (Overton, 2006a), conduz a ponderar as implicações teóricas e metateóricas dos resultados da presente investigação.

[15] Ao tomar o governo das nações como metáfora da inteligência humana, na Teoria do Auto-Governo Mental (ver CAPÍTULO 3, pp.205-206), Sternberg dá lugar, ainda que apenas de forma implícita, à ideia de que a inteligência humana pode ser tomada como metáfora do governo das nações.

9.2.IMPLICAÇÕES TEÓRICAS DOS RESULTADOS

9.2.1. *WAIS-III* (2002) + *STAT-R(H)* (2004):

Síntese e Significado Teórico dos Resultados

Apesar de necessariamente circunscritos, e de condicionados por algumas importantes limitações metodológicas e metrológicas oportunamente assinaladas, os resultados desta investigação terão tido pelo menos o mérito de estimular a reflexão sobre a natureza e o lugar de algumas teorias da inteligência humana no quadro dos grandes paradigmas de investigação e, mais fundamentalmente, sobre as metateorias e as visões ontológicas do mundo em que se inspiram e que, em última análise, representam. No CAPÍTULO 2 ficaram apontados alguns eixos organizadores desta reflexão, os quais serão agora retomados em articulação não só com os resultados obtidos, mas ainda com as considerações que suscitaram, relativas à avaliação da inteligência, abordadas no tópico precedente.

As teorias factoriais de organização das aptidões (ver CAPÍTULO 1) constituem a mais característica conceptualização diferencial da inteligência humana. Podendo assumir várias formas, muitas vezes decorrentes dos postulados de que partem, muitas vezes decorrentes (também) das opções metodológicas de que dependem, estas teorias construíram-se a partir da evidência das diferenças individuais na eficiência de resolução de problemas suficientemente complexos para adquirirem valor preditivo em relação a critérios relevantes do sucesso dos indivíduos em contextos significativos de vida – designadamente, a escola e a vida profissional. O "sucesso", definido objectivamente, reportou-se por muito tempo a critérios externos, de avaliação escolar ou de desempenho profissional, pelo que as técnicas diferenciais, capazes de predizer esses critérios, se mostravam "válidas" para essa finalidade. Mas estas teorias factoriais, mesmo as mais complexas que se estruturam numa hierarquia contemplando vários "níveis de generalidade", organizam a variabilidade inter-individual que se manifesta num nível, entre outros conceptualizáveis, de observação e explicação – o do desempenho nas tarefas que compõem as técnicas diferenciais das aptidões. Por isso, os diversos modelos factoriais representam formas alternativas mas equivalentes, nenhuma "melhor" do que outra, de "explicar" a variância inter-individual num determinado nível de observação e explicação; e porque todos os modelos acabam por apresentar potencialidades, os hieráquicos, como a Teoria dos Três Estratos de Carroll ou o Modelo *C-H-C* (Cattell-Horn-Carroll) (Carroll, 2003, p.18), exactamente por serem mais completos e integrativos (de novo, num determinado nível de observação), ganharam aceitação consensual e constituíram-se, até hoje, como o quadro de referência para a avaliação diferencial da cognição humana. Este tipo de modelo factorial tem atrás de si não só décadas de investigação factorial, como um dos mais impressionantes estudos de meta-análise alguma vez efectuado em psicologia, publicado em 1993 por John Carroll, envolvendo 60 anos de investigação

conduzida nos cinco continentes, com enorme diversidade de instrumentos e de amostras, num total de quase 500 conjuntos de dados – uma investigação impar, sem dúvida, ainda que extemporânea, na opinião de Sternberg (1994a), perante os contributos então emergentes no domínio da inteligência.

Com efeito, pela altura desta publicação já a psicologia da inteligência tinha trilhado novos caminhos e alargado muito significativamente os seus horizontes. Emergiam, então, as perspectivas sistémicas de investigação e de avaliação da inteligência: em 1983, Gardner apresentara uma teoria que, ao pluralizar o termo "inteligência", acentuava o carácter limitado do conceito tradicional; em 1985 Sternberg publicara a sua obra *Beyond IQ*, um título que desde logo deixava transparecer a sua posição crítica relativamente à natureza reducionista do conceito tradicional de inteligência; em 1990 Salovey e Mayer haviam introduzido o novo conceito de Inteligência Emocional, que também sublinhava, só por si, o reconhecimento da limitação do espectro de funcionamento abrangido pela noção tradicional de inteligência; e no mesmo ano Ceci publicara a primeira edição da obra *On intelligence,* onde apresentava a Teoria Bioecológica da INTELIGÊNCIA humana. Um traço comum a todas estas propostas sistémicas de investigação da inteligência: a inclusão nos modelos da inteligência de outros níveis de observação e explicação, para lá do nível do desempenho nos testes convencionais, através do alargamento do construto tendo em vista a inclusão da interacção com o contexto, ampliando, desse modo, o âmbito das predições que viabiliza.

Os fundamentos e contributos teóricos destes dois paradigmas de investigação da inteligência – o diferencial e o sistémico – foram explorados nos capítulos de enquadramento teórico (CAPÍTULOS 1 a 3); uma técnica de avaliação representativa de cada um foi traduzida, adaptada e aperfeiçoada (CAPÍTULOS 4 a 7), tendo em vista prepará-la para o estudo da estrutura das medidas emergente e do grau em que esta esclarece sobre a natureza dos construtos e as suas inter-relações; e os resultados deste estudo foram apresentados e discutidos (CAPÍTULO 8). Pretende-se agora analisar, para lá das implicações dos resultados deste estudo para a medida da inteligência com recurso a essas duas técnicas (CAPÍTULO 9, TÓPICO 9.1), também as implicações destes resultados do ponto de vista das teorias e das metateorias que fundamentam essas medidas. Para isso, parte-se da síntese dos principais resultados do estudo conjunto das duas técnicas (*WAIS-III* e *STAT-R (H)*):

- Importa começar por recordar que os estudos correlacionais se mostraram francamente condicionados pelas limitações quanto à consistência interna de muitas medidas do *STAT-R (H)* (de onde resultaram, por exemplo, intercorrelações mais baixas no *STAT-R (H)* do que na *WAIS-III*; nas análises factoriais exploratórias, menores comunalidades (h^2) no *STAT-R (H)* do que na *WAIS-III;* e no estudo da dimensionalidade do *STAT-R (H)*, muito marcada presença de ruído nos dados, ou seja, modelo explicando em geral proporção limitada da variância);

mesmo assim, e uma vez verificadas as condições mínimas para a aplicação dos métodos de análise factorial e de análise dimensional, foi possível encontrar algumas tendências de resultados com interesse, do ponto de vista do teste das Hipóteses Experimentais estabelecidas no CAPÍTULO 4;

- Na análise da matriz de intercorrelações conjunta, verificam-se correlações apreciáveis entre medidas das duas técnicas de avaliação utilizadas, sugerindo que tendem a medir construtos comuns; particularmente importantes são as correlações entre Aritmética (*WAIS-III*) e a Parte 2, Analítica/Quantitativa (*STAT-R (H)*) e entre Cubos e Matrizes (*WAIS-III*) e as Partes 3 e 8, Analítica/Figurativa e Criativa/Quantitativa (*STAT-R (H)*); ainda na análise das intercorrelações, manifestam-se de interesse as correlações entre todas as partes Criativas (*STAT-R (H)*) e os subtestes de Cubos e Matrizes;
- Nas intercorrelações de resultados compósitos foi importante verificar que todos os resultados do *STAT-R (H),* relativos quer aos domínios de processamento, quer às áreas de conteúdo, se correlacionam muito significativamente com o QI de Escala Completa da *WAIS-III* (.50 a .71) (cf. QUADRO 8.4), embora de maneira mais expressiva o resultado Criativo e o resultado Quantitativo; e ainda, que o resultado do domínio Criativo se correlaciona elevadamente com o Índice de Organização Perceptiva, e com a parte de Realização, e o domínio Analítico com o índice de Memória de Trabalho e com a parte Verbal;
- O resultado mais consistentemente replicado ao longo das várias análises consistiu na identificação de uma componente comum a todas as variáveis envolvidas nos vários níveis de análise: 1) as matrizes factoriais antes da rotação, quer na análise em componentes principais de cada instrumento, quer na análise factorial exploratória dos dois instrumentos em conjunto, apresentam uma componente ou um factor que explica uma parte substancial da variância dos resultados e em que saturam positivamente todas as variáveis de partida; 2) as análises da dimensionalidade do conjunto de todos os itens ou de grupos seleccionados de itens, puseram em evidência o seu ajustamento global, e de uma larga maioria dos sujeitos, à dimensão do modelo, tendo em geral ficado por explicar uma parte diminuta da variância dos resultados; acresce que todos os domínios de processamento, indistintamente, (não só o domínio Analítico), e todas as áreas de conteúdo, são compostos por itens que se ajustam ao modelo definido a partir dos itens da *WAIS-III* e do *STAT-R (H)* em conjunto; 3) por fim, o estudo

confirmatório das medidas, embora não tenha demonstrado como mais ajustado o modelo de Factor Geral, admitiu um maior ajustamento de modelos com factores oblíquos, de onde decorre o reconhecimento implícito da presença de um factor comum aos factores de primeira ordem (os modelos de factores ortogonais foram os que alcançaram índices de ajustamento mais reduzidos);

- Quer as análises factoriais quer o estudo da dimensionalidade tenderam a fazer emergir dois a três factores, para além do factor geral:
  - Na análise factorial exploratória: após rotação dos eixos, FACTOR I – Subtestes de Realização (saturações mais altas: Cubos e Matrizes) (*WAIS-III*) e Partes 8 (Criativa/Quantitativa), 3 e 9 (Analítica/Figurativa e Criativa/Figurativa, ambas provas de matrizes) (*STAT-R (H)*); FACTOR II – Subtestes de Vocabulário, Informação, Semelhanças e Compreensão (*WAIS-III*) e nenhuma saturação significativa do *STAT-R (H)*; FACTOR III – Subteste de Aritmética (e saturação mais elevada de Memória de Dígitos) (*WAIS-III*) e Partes 2 e 5 Analítica/Quantitativa e Prática/Quantitativa (*STAT-R (H)*); o primeiro factor pode ser interpretado como INTELIGÊNCIA FLUIDA, o segundo como INTELIGÊNCIA CRISTALIZADA e o terceiro como NUMÉRICO/MEMÓRIA DE TRABALHO.
  - Na análise dos contrastes: nas análises em que entraram todos os itens da *WAIS-III* e itens do *STAT-R (H)* (todos os itens ou apenas os de cada domínio ou de cada área), a percentagem de variância explicada pelo modelo foi sempre muito elevada, muito mais elevada do que a proporção explicada pelo 1º contraste; ainda assim, este contraste, sempre definido a partir de itens ajustados ao modelo (o que significa que não corresponde a uma dimensão distinta da representada pelo modelo mas antes a um efeito local na dimensão do modelo), definiu-se pelas saturações opostas na *WAIS-III*, entre itens de Informação e Vocabulário e itens de Cubos e Matrizes; este resultado parece sugerir a distinção entre INTELIGÊNCIA CRISTALIZADA e INTELIGÊNCIA FLUIDA, não a distinção entre domínios de processamento, sugerida pela Teoria Triárquica; esta aparece apenas vagamente esboçada na análise em que se trataram em conjunto todos os itens do *STAT-R (H)* e os itens de Compreensão Verbal da *WAIS-III* na qual o 1º contraste se definiu pela oposição entre as saturações de itens dos subtestes de Informação e Vocabulário (*WAIS-III*) e das Partes Criativa e Prática (embora também alguns da Analítica) (*STAT-R (H)*).

- Na análise factorial confirmatória: os modelos mais ajustados à estrutura empírica das variáveis (Modelos 4d e 5b) foram os que se estruturaram em três factores oblíquos – Compreensão Verbal, Organização Perceptiva e Numérico/Memória de Trabalho – mas admitindo a saturação da Parte 8 (Criativa/Quantitativa) do *STAT-R (H)* no factor de Organização Perceptiva, uma solução inspirada nas intercorrelações e nos resultados da análise factorial exploratória; uma possível leitura desta organização das medidas remete de novo para um factor de INTELIGÊNCIA CRISTALIZADA, um factor de INTELIGÊNCIA FLUIDA e um factor NUMÉRICO/MEMÓRIA DE TRABALHO.

- Por fim, importa ainda reter nesta síntese que a hipótese experimental que obteve maior apoio foi a HIPÓTESE EXPERIMENTAL 4.2 – inspirada na estrutura hierárquica clássica das aptidões humanas – não saindo apoiadas as Hipóteses Experimentais 2 e 4.1, que representavam a Teoria Triárquica, nem a Hipótese 3 (multifactorial) ou a Hipótese 1 (factor geral); assinale-se, contudo, que a Hipótese 4.2 não exclui a presença de g nos resultados, apenas não lhe atribui o estatuto de única dimensão subjacente à variância dos resultados obtidos com a *WAIS-III* e o *STAT-R (H)* na amostra do Estudo Principal.

A mais aceite concepção estrutural das aptidões humanas, representada pelo modelo hoje abreviadamente designado *C-H-C*, e que acaba por sintetizar as posições de autores tão diferentes como Spearman, Thurstone (na fase mais avançada dos seus trabalhos), Cattell e Carroll, consiste em reconhecer a existência de uma factor de inteligência geral ou *g*, que pode ser empiricamente confirmado, juntamente com cerca de dez factores amplos de segundo estrato, entre os quais se destacam os factores designados como *Gf* e *Gc* (inteligência fluida e inteligência cristalizada), associados a tipos específicos de variáveis (Nyborg, 2003, p.3). Esta concepção de organização das aptidões apresenta-se em larga medida coerente com a concepção de inteligência de Wechsler, e até com a estrutura básica das suas escalas – por um lado, pelo destaque atribuído ao factor geral, ou seja, a uma inteligência global necessária a todo o desempenho cognitivo; por outro lado, porque a distinção que introduziu nas suas provas, entre duas partes, Verbal e Realização, deixa transparecer, apesar da noção global de inteligência, o reconhecimento da utilidade de resultados de nível de generalidade inferior (QI Verbal e QI de Realização) referidos a gamas de funcionamento cognitivo mais restritas: um resultado envolvendo subtestes que, ao apelar ao uso da linguagem e a conhecimentos adquiridos formal e informalmente em função da experiência, sugere proximidade à noção de inteligência cristalizada (*Gc*) e o outro envolvendo subtestes pouco dependentes do uso da linguagem e que exigem a resolução de problemas novos, que apelam pouco a aprendizagens

anteriores, alguns de conteúdo abstracto (*Gf*). Os resultados obtidos na presente investigação mostram-se francamente coerentes com esta perspectiva da organização das aptidões, de onde decorre, como primeira implicação teórica, a confirmação da concepção hierárquica das aptidões humanas na base das medidas da *WAIS-III* e do *STAT-R (H).*

Mas esta implicação põe em causa uma outra concepção, a que deriva da Teoria Triárquica da Inteligência: de acordo com Sternberg, esperar-se-ia, como se viu antes, uma organização do conjunto de todas as variáveis medidas pela *WAIS-III* e pelo *STAT-R (H)* em torno de três domínios de processamento – Analítico, Prático e Criativo. E, embora esses domínios possam envolver sempre as mesmas componentes de processamento da informação (Subteoria Componencial) estas podem aplicar-se em situações mais familiares ou menos familiares – Inteligência Analítica *versus* Inteligência Criativa (Subteoria Experiencial), ou lidar com conteúdos mais abstractos, semelhantes aos que envolvem as tarefas escolares, ou mais concretos, semelhantes aos que envolvem as situações da vida comum – Inteligência Analítica *versus* Inteligência Prática (Subteoria Contextual)[16]. Os resultados do presente estudo, há que reconhecê-lo, não apoiam esta forma de conceber a inteligência humana, pelo menos a partir das operacionalizações adoptadas pelos dois autores, o que constitui uma segunda implicação teórica deste estudo: primeiro, a estrutura factorial emergente da análise exploratória não se configura de acordo com o esperado com base na Teoria Triárquica – posto que não emergem os três domínios de processamento, com o domínio Analítico a saturar mais fortemente no factor em que saturam as medidas proporcionadas por uma técnica convencional de avaliação da inteligência (*WAIS-III*); e estes domínios não emergem nem como dimensões ortogonais, como postula a teoria à partida, nem sequer como dimensões oblíquas, uma vez que os modelos baseados na separação destes três domínios não foram os mais ajustados aos dados na análise confirmatória; segundo, a análise dimensional mostrou que a grande maioria dos itens do *STAT-R (H)* se ajustam a uma dimensão única definida pelos itens da *WAIS-III* e que o primeiro contraste, em todas as análises em que a totalidade dos itens da *WAIS-III* entrem, delimita uma dimensão dentro do modelo que pode ser interpretada a partir do contraste entre Inteligência Cristalizada e Inteligência Fluida – não como uma distinção entre as dimensões Analítica e Prática ou Analítica e Criativa; terceiro, os modelos que a análise confirmatória demonstrou mais ajustados à estrutura empírica das variáveis não são os que se baseiam nos postulados da Teoria Triárquica, mas os que se baseiam numa estrutura mais próxima da clássica concepção hierárquica das aptidões humanas.

Uma terceira implicação teórica dos resultados assenta numa verificação empírica deste estudo que corrobora resultados obtidos em análises conjuntas do *STAT-R (H)* com outras medidas

[16] Note-se que o autor é bem menos explícito quanto à distinção entre Inteligência Prática e Inteligência Criativa, mas presume-se que a primeira envolva situações mais familiares do que a segunda, posto que envolve os problemas enfrentados no quotidiano dos contextos comuns de vida do indivíduo.

tradicionais da inteligência: a de que as partes Criativas do *STAT-R (H)* se ligam mais claramente a medidas da inteligência geral (QI de Escala Completa da *WAIS-III*, por exemplo), ou a medidas da inteligência fluida (subtestes de Matrizes e de Cubos, factor de Organização Perceptiva, parte de Realização da *WAIS-III*, por exemplo), do que as partes Analíticas, como faria supor tanto a Teoria Triárquica, como a critica insistente de Sternberg ao carácter limitado da gama de funcionamento cognitivo abrangida pelos testes tradicionais, a qual não contempla a Inteligência Criativa. Com efeito, é oportuno recordar que o conceito de inteligência fluida comporta, quanto mais não seja ao ser demarcado do de inteligência cristalizada, a noção de que envolve a resolução de tarefas novas, com as quais o indivíduo não se confrontou no passado e para as quais não desenvolveu, por consequência, conhecimento organizado ou automatismos. Não será surpreendente, por consequência, que as medidas Criativas do *STAT-R (H)* se correlacionem com os testes de inteligência fluida como o de Matrizes e o de Cubos, ou saturem em factores que envolvem várias medidas da inteligência fluida, como o de Organização Perceptiva e a parte de Realização da *WAIS-III*, ou ainda em testes de Factor *g* (Sternberg, 2003f).

Importa também notar que a "novidade" é introduzida nas partes Criativas do *STAT-R (H)* ao nível dos conteúdos - itens verbais com conteúdo contrafactual, condições "novas" a verificar antes de efectuar cálculos aritméticos, ou matrizes figurativas distintas para extracção e para aplicação da regra de resolução – e não ao nível das tarefas, que são até relativamente comuns nos testes de inteligência – analogias verbais, operações aritméticas e matrizes figurativas. Como foi antes assinalado, esta opção mostra-se algo paradoxal, uma vez que o autor considera as operações ou processos mentais como mais fundamentais ao funcionamento cognitivo do que os conteúdos a que se aplicam esses processos, pelo que seria de esperar que tivesse proposto *tarefas* pouco familiares, e não simplesmente *conteúdos* (supostamente) menos familiares. Acresce que a resposta a analogias verbais, o cálculo aritmético e a resolução de matrizes figurativas são consideradas pelo autor tarefas representativas da inteligência Analítica, uma vez que são bem tradicionais em testes de inteligência, razão por que surpreende que sejam retomadas na construção de uma medida *nova* da Inteligência Criativa. Ao estabelecer o contraste em função do grau de familiaridade das tarefas, do qual decorre a distinção entre Inteligência Analítica e Inteligência Criativa, Sternberg parece retomar, afinal, a distinção, fundamental e já bem estabelecida em psicologia da inteligência, entre Inteligência Cristalizada e Inteligência Fluida.

Um resultado mais consistente com a Teoria Triárquica diz respeito à Inteligência Prática: repetidamente, em todas as análises, o domínio Prático apresenta menor saturação no factor de inteligência geral definido a partir das medidas conjuntas da *WAIS-III* e do *STAT-R (H)* demonstrando poder ser entendido como um construto algo separado das noções precedentes – de Inteligência

Analítica e Inteligência Criativa ou de Inteligência Cristalizada e Inteligência Fluida. Esta constatação é coerente com a observação de Sternberg e colaboradores (ver, por exemplo, Sternberg, Ferrari, Clinkenbeard & Grigorenko, 1996) relativamente a que as partes práticas do *STAT-R (H)* tendem a ser as que apresentam correlações mais baixas com os testes convencionais de inteligência. Mesmo assim, na presente investigação esta distinção está longe de ser suficientemente clara ou inequívoca para poder ser considerado que apoia as predições da Teoria Triárquica de forma expressiva.

Uma vez que as constatações empíricas tenderam a confirmar mais as expectativas que se poderiam estabelecer com base no modelo tradicional de organização das aptidões do que as que poderiam derivar da Teoria Triárquica, esta falta de correspondência entre os resultados do *STAT-R (H)* e a teoria pode ser interpretada genericamente de dois modos: ou o *STAT-R (H)* constitui uma operacionalização deficiente da Teoria Triárquica, e então não será legítimo utilizá-lo, mesmo que apenas em contextos de investigação, para obter medidas da Inteligência Funcional; ou o que mais fundamentalmente pode ser posto em causa é a própria Teoria Triárquica, como modelo teórico da inteligência. Quanto à primeira interpretação, o presente estudo não deixou de assinalar múltiplas limitações metrológicas, em boa parte confirmadas em estudos internacionais, as quais se reflectem no rigor e no significado das medidas que o *STAT-R (H)* proporciona. Muito em particular, o facto de os estudos da estrutura das medidas não confirmarem de forma clara a configuração de variáveis teoricamente esperada tem como implicação que não seja legítimo interpretar os resultados das medidas remetendo para os construtos de Inteligência Analítica, Inteligência Prática e Inteligência Criativa, e, mais importante, que não seja sequer legítimo efectuar os somatórios respectivos de pontuações dos itens, posto que não se organizam em medidas lineares separadas.

Mas a segunda interpretação torna-se neste momento mais pertinente, ao situar-se mais claramente no plano teórico que se pretende discutir: poderá a própria Teoria Triárquica ser questionada enquanto modelo teórico da inteligência? Certamente não com base num único estudo, não com base em umas quantas amostras e mais ainda com base numa versão traduzida do *STAT-R (H)*. Não se pretende de todo pôr em causa todo o corpo de investigação que tem vindo a ser desenvolvido desde há mais de dez anos por Sternberg e seus colaboradores com o propósito de proceder à validação da Teoria Triárquica, e nem se considera dispor de suficiente fundamento a partir dos dados limitados do presente estudo. Mas estes dados sugeriram algumas reflexões que se admite terem implicações para a investigação e a medição da inteligência, uma vez que dizem respeito aos fundamentos epistemológicos e metateóricos da investigação da inteligência humana.

A persistente verificação da organização das variáveis do *STAT-R (H)* em dimensões comuns às medidas tradicionais da inteligência, e designadamente, neste estudo, à *WAIS-III*, leva a considerar que possam constituir apenas novas formas de medir a mesma inteligência, uma noção coincidente,

afinal, com a concepção de inteligência de Wechsler. Mas esta verificação pode significar, tão só, que as medidas do *STAT-R (H)* se situam no mesmo nível de observação e explicação em que se situam as medidas convencionais de inteligência, ao reportarem-se, afinal, a aspectos do mundo interno do indivíduo. Como foi assinalado a propósito do *STAT-R (H)* (TÓPICO 9.1 deste capítulo), muitas das suas características aproximam-no das técnicas tradicionais colectivas de medida da inteligência, ao mesmo tempo que estão ausentes as características que permitiriam uma aproximação a uma avaliação contextualista – flexível porque adaptativa, dinâmica e baseada na relação entre factores conativos, afectivos e cognitivos. A questão é agora recolocada no plano teórico: será que a própria Teoria Triárquica, apesar de incorporar uma subteoria contextual, que sublinha que a inteligência se define em função do contexto, e uma subteoria experiencial, que sublinha que a inteligência se define também em função da experiência e das aprendizagens passadas, acaba por perder o seu carácter sistémico ao equacionar essas subteorias como dimensões – Inteligência Prática e Inteligência Criativa – por inerência "quantificáveis" [17]? E será que a Teoria Triárquica (ou pelo menos a sua operacionalização no *STAT-R (H)*) permanece, por isso, situada num nível de observação relativamente restrito do contínuo que pode ser conceptualizado entre as variáveis moleculares do mundo interno do indivíduo (objecto do paradigma biológico) e as variáveis molares do mundo externo ou do contexto (objecto dos paradigmas sociológico e antropológico)? Estas questões equivalem, afinal, a questionar a sua natureza "sistémica".

A noção de medida e a noção de dimensão são difíceis de separar, porque impossíveis de conceber separadamente: a medida pressupõe variação quantitativa num contínuo; e é, por definição, a variabilidade inter-individual no contínuo que faz emergir uma dimensão. Já em 1965, na terceira edição da obra *The Psychology of Human Diffrences*, Leona Tyler[18] (1906-1993), ao efectuar um balanço da Psicologia Diferencial até à época, e antecipar as tendências de evolução para o final do século, acentuava o contraste entre a perspectiva clássica do domínio, nomotética, centrada na avaliação quantitativa ou medida e na noção de dimensão, e a perspectiva que iria crescer em importância nos anos seguintes, idiográfica, centrada na avaliação qualitativa do desenvolvimento e em novos conceitos mais flexíveis, porque multifacetados (não necessariamente unidimensionais) e contextualistas – como estilos, estratégias, tipos (definidos como direcções tendenciais de desenvolvimento), variáveis moderadoras, etc. A noção de estilo, por exemplo, é contextualista na medida em que não se define pela quantidade de um atributo (como uma aptidão ou um traço no sentido clássico) mas pelos processos mentais predominantemente escolhidos, que podem variar em

[17] Um exemplo particularmente ilustrativo do carácter dimensional dos conceitos da Teoria Triáqruica da Inteligência Funcional encontra-se na análise das suas implicações para as escolhas vocacionais e para o desenvolvimento da carreira (Sternberg, 2003e), na qual as inteligências analítica, prática e criativa são tratadas como variáveis "de conteúdo", não como variáveis de "processo" (Afonso, 2005a, 2006a).

[18] A quarta mulher a tornar-se presidente da Associação Americana de Psicologia, em 1972/73.

função das situações, tornando o comportamento mais ou menos adaptado ao contexto (Gilles, 1999), uma noção que invoca o conceito de processo vicariante, mais tarde introduzido por Reuchlin (1978a; 1988, 1999/2002). Há uma evidente mudança de visão do mundo na transição da perspectiva dimensional e centrada na medição (formista/mecanicista) para a perspectiva desenvolvimentista e centrada na descrição (organicista/contextualista). Mais, a convicção mais básica de Tyler era na realidade inspirada numa visão do mundo selectivista: "o desenvolvimento pessoal depende da selecção e actualização, a partir de uma infinidade de possibilidades, daquelas poucas que o tempo e a situação viabilizam" (Littman, 1993). Equacionar na forma dimensional qualquer conceito, com o intuito de o tornar mensurável, traduz uma opção implícita por uma visão do mundo formista ou mecanicista, logo pouco compatível com qualquer quadro de referência teórico que se pretenda sistémico; esta incoerência é ainda mais acentuada se esse quadro de referência remeter para a noção de sistema altamente complexo, caracterizado pela consciência das suas próprias finalidades e pela intencionalidade, a qual traduz uma opção implícita por uma visão do mundo selectivista.

Será talvez oportuno notar, como o faz Van Geert (2003b), que os modelos factoriais, que organizam as dimensões psicológicas pertinentes à avaliação das diferenças individuais – aptidões, inteligência, etc. – são modelos que reportam a maneira como essas dimensões variam *entre* os indivíduos, escolhidos enquanto membros de determinado grupo ou população; mas não existe nenhuma relação lógica ou conceptual entre a estrutura factorial da inteligência definida a partir das diferenças *entre* os indivíduos e a estrutura da inteligência *dentro* dos indivíduos. E a razão é simples: a estrutura da inteligência em cada indivíduo é o produto de um processo dinâmico e temporal enquanto a estrutura da inteligência num grupo é produto das diferenças entre indivíduos independentes. Daqui decorre sem dúvida a acusação clássica dirigida aos modelos factoriais, por constituírem descrições "estáticas" que não tomam em consideração a variabilidade intra-individual, em função do tempo e da situação; como decorre também a natureza pouco contextualizada da avaliação psicológica tradicional.

Esta limitação dos modelos factoriais fica equacionada com clareza ao retomar a tabela inspirada em Stern (1911 citado em Gilles, 1999) e apresentada no CAPÍTULO 4 (TABELA 4.1, p.252): a análise factorial (à semelhança de muitos outros métodos de análise multivariada, como a modelização estrutural ou a modelização de Rasch, acrescente-se) consiste numa metodologia que prescinde, regra geral, de considerações temporais ou situacionais – as medidas de que parte são supostamente independentes do momento da avaliação (procura-se averiguar em que grau pelo estudo da estabilidade temporal das medidas) e são obtidas em situações ou condições tanto quanto possível uniformes (estandardização)[19]. Ora, a inclusão das instâncias temporal e situacional no esquema

[19] Será pertinente relembrar que Raymond Cattell, no domínio conativo, identificou e aplicou diversos planos factoriais, alguns dos quais envolviam a intervenção de variáveis temporais (Plano P, baseado no estudo de um único caso observado

clássico de Stern produz uma desmultiplicação dos planos de experiência que os aproxima de uma abordagem metodológica de natureza sistémica, e muito em particular da abordagem dos sistemas dinâmicos, adiante caracterizada (CAPÍTULO 10). Por outras palavras, a abordagem diferencial clássica, porque baseada no postulado da estabilidade temporal das diferenças individuais e na natureza nomotética da definição e da medição das dimensões das diferenças (com ênfase nas diferenças inter-individuais e inter-grupais), omitiu intencionalmente dos seus planos de experiência a consideração das instâncias temporal e contextual: a primeira, pelo acento na consistência comportamental e na aplicação de métodos de precisão "teste-reteste"; a segunda, pelo acento no controlo das condições de observação e pela aposta na "estandardização" da medição. De fora deixou, por muito tempo, a variabilidade intra-individual e a variabilidade inter-situacional, o que a afastou da psicologia experimental, tal como denunciou Cronbach em 1957.

A variabilidade intra-individual em curto intervalo de tempo é em geral lida pela Teoria Clássica dos Testes como resultante de erro de medida, não como produto de uma interacção dinâmica e contínua entre o indivíduo em desenvolvimento e o contexto em mudança. Como a história da Psicologia bem demonstrou, entender as diferenças como erro conduz a perspectivas simplificadoras dos fenómenos e enraizadas em visões do mundo limitadas (recorde-se a posição dos primeiros psicólogos experimentais relativamente às diferenças individuais, tratadas como erro de observação); deste ponto de vista a psicologia diferencial por muito tempo teve em relação às diferenças intra-individuais e inter-situacionais a mesma atitude que os primeiros psicólogos experimentais tiveram em relação às diferenças inter-individuais: remeteu-as para a variância de erro. Entender as diferenças *inter* indivíduos como objecto de estudo, fez emergir, no virar do século XIX para o XX, a nova psicologia ocupada das diferenças *inter*-individuais e *inter*-grupais; mas entender as diferenças *intra* indivíduos como objecto de estudo, uma marca distintiva desse mesmo domínio diferencial na actualidade, implica assumir incontornavelmente a inteligência no quadro dos "sistemas dinâmicos", que contemplam, por definição, as instâncias temporal e situacional, e ultrapassar as visões do mundo redutoras dos fenómenos a classificações taxonómicas (como as "estruturas das aptidões") ou a mecanismos mentais (como os algoritmos de "processamento mental da informação"). No plano metateórico, implica renunciar à metateoria fragmentada, que simplifica os fenómenos ao postular dimensões discretas, lineares, independentes e quantificáveis, ao admitir o seu carácter aditivo e ao aceitar para a medida o enquadramento na Teoria Clássica dos Testes; e antes assumir a metateoria relacional (Overton, 2006a; Overton & Ennis, 2006), que equaciona a mudança, a interacção e a

em múltiplas ocasiões nas mesmas variáveis). Mas o plano de análise mais utilizado no domínio cognitivo envolve as intercorrelações de variáveis medidas numa única ocasião numa amostra de indivíduos (Plano R) (Anastasi, 1958, p.389).

complexidade como objectos de estudo e concebe a avaliação muito para lá da simples quantificação de atributos.

Nesta perspectiva, as limitações metrológicas encontradas no *STAT-R (H),* quer quanto à precisão quer quanto à estrutura interna das medidas, que podem bem ser lidas sumariamente como resultado de deficiente operacionalização da teoria, sugerem antes outras leituras mais profundas: uma, a de que não existe coerência, do ponto de vista metateórico e epistemológico, entre a Teoria Triárquica e a sua operacionalização, o *STAT-R (H)*: a teoria, como se viu no CAPÍTULO 3, evoluiu de um enquadramento mecanicista, evidente na primeira fase de investigação dominada pela análise componencial (Sternberg, 1977), para um enquadramento contextualista, já patente na obra *Beyond IQ* de 1985, e mais tarde para um enquadramento selectivista, a partir da introdução, em 1996, do conceito de Inteligência Funcional; a medição, como se viu atrás, enquadra-se, por contraste, na abordagem diferencial clássica, assente numa visão do mundo quanto muito mecanicista, ao inferir o funcionamento de mecanismos mentais subjacentes ao comportamento observado em situação estandardizada, estática e, na verdade (apesar da subteoria contextual), descontextualizada.

Mas outra leitura possível, ainda mais ousada sem dúvida, é a de que a própria Teoria Triárquica, ao desenvolver-se no sentido de favorecer a operacionalização em técnicas de avaliação diferencial e a validação com recurso a métodos enquadrados na Teoria Clássica dos Testes, renunciou à natureza sistémica que almejava e acabou por restringir o nível de observação e explicação em que se posiciona àquele para que necessariamente remete a *avaliação quantitativa* de *dimensões psicológicas*. Neste sentido, a transição entre a primeira formulação da "Teoria Triárquica" (1985) e a actual "Teoria Trárquica da Inteligência Funcional" (uma designação cada vez mais frequente nos trabalhos do seu autor) terá tornado o modelo mais inteligível, mais acessível à operacionalização em técnicas diferenciais (e também, há que dizê-lo, mais passível de alcançar popularidade entre não especialistas), mas em parte à custa de aproximar os conceitos postulados na teoria de noções dimensionais, tanto mais intuitivas quanto enquadradas numa lógica mecanicista, bem tradicional no domínio da inteligência e bem implantada no senso comum. Assim, as subteorias "Componencial", "Experiencial" e "Contextual", da Teoria Trárquica original, sem serem abandonadas reorganizam-se em três conceitos "arriscados", da Teoria Triárquica da Inteligência Funcional – "Inteligência Analítica", "Inteligência Criativa" e "Inteligência Prática" – "arriscados" por incorrerem no risco de retomar a tradicional definição dimensional da inteligência e o respectivo formato de avaliação, ou seja, o mesmo nível de observação e explicação de há muito (o que fica em evidência nos resultados dos estudos factoriais e dimensionais onde as medidas triárquicas emergem apenas como novas medidas da "mesma" inteligência). E "arriscados", por isso mesmo, por sacrificarem, afinal, o propósito original de construção de uma teoria sistémica da inteligência humana.